

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/49, A61K 48/00		A2	(11) International Publication Number: WO 00/39302 (43) International Publication Date: 6 July 2000 (06.07.00)
(21) International Application Number: PCT/US99/31245		(72) Inventors: BARNETT, Susan; Chiron Corporation, 4560 Horton Street – R440, Emeryville, CA 94608 (US). ZUR MEGEDE, Jan; Chiron Corporation, 4560 Horton Street – R440, Emeryville, CA 94608 (US). SRIVASTAVA, Indresh; Chiron Corporation, 4560 Horton Street – R440, Emeryville, CA 94608 (US). LIAN, Ying; Chiron Corporation, 4560 Horton Street – R440, Emeryville, CA 94608 (US). HARTOG, Karin; Chiron Corporation, 4560 Horton Street – R440, Emeryville, CA 94608 (US). LIU, Hong; Chiron Corporation, 4560 Horton Street – R440, Emeryville, CA 94608 (US). GREER, Catherine; Chiron Corporation, 4560 Horton Street – R440, Emeryville, CA 94608 (US). SELBY, Mark; Chiron Corporation, 4560 Horton Street – R440, Emeryville, CA 94608 (US). WALKER, Christopher; Chiron Corporation, 4560 Horton Street – R440, Emeryville, CA 94608 (US).	
(22) International Filing Date: 30 December 1999 (30.12.99)		(74) Agents: DOLLARD, Anne, S.; Chiron Corporation, Intellectual Property – R440, P.O. Box 8097, Emeryville, CA 94662–8097 (US) et al.	
(30) Priority Data: 60/114,495 31 December 1998 (31.12.98) US 60/168,471 1 December 1999 (01.12.99) US		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
Published <i>Without international search report and to be republished upon receipt of that report.</i>			
(54) Title: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION OF VIRUS-LIKE PARTICLES			
(57) Abstract			
<p>The present invention relates to the efficient expression of HIV polypeptides in a variety of cell types, including, but not limited to, mammalian, insect, and plant cells. Synthetic expression cassettes encoding the HIV Gag-containing polypeptides are described, as are uses of the expression cassettes in applications including DNA immunization, generation of packaging cell lines, and production of Env-, tat- or Gag-containing proteins. The invention provides methods of producing Virus-Like Particles (VLPs), as well as, uses of the VLPs including, but not limited to, vehicles for the presentation of antigens and stimulation of immune response in subjects to whom the VLPs are administered.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND
PRODUCTION OF VIRUS-LIKE PARTICLES

5 **TECHNICAL FIELD**

Synthetic expression cassettes encoding the HIV polypeptides (e.g., Gag-, pol-, prot-, reverse transcriptase, Env- or tat-containing polypeptides) are described, as are uses of the expression cassettes. The 10 present invention relates to the efficient expression of HIV polypeptides in a variety of cell types. Further, the invention provides methods of producing Virus-Like Particles (VLPs), as well as, uses of the VLPs and high level expression of oligomeric envelope proteins.

15

BACKGROUND OF THE INVENTION

Acquired immune deficiency syndrome (AIDS) is recognized as one of the greatest health threats facing modern medicine. There is, as yet, no cure for this 20 disease.

In 1983-1984, three groups independently identified the suspected etiological agent of AIDS. See, e.g., Barre-Sinoussi et al. (1983) Science 220:868-871; Montagnier et al., in Human T-Cell Leukemia Viruses 25 (Gallo, Essex & Gross, eds., 1984); Vilmer et al. (1984) The Lancet 1:753; Popovic et al. (1984) Science 224:497-500; Levy et al. (1984) Science 225:840-842. These isolates were variously called lymphadenopathy-associated virus (LAV), human T-cell lymphotropic virus

type III (HTLV-III), or AIDS-associated retrovirus (ARV). All of these isolates are strains of the same virus, and were later collectively named Human Immunodeficiency Virus (HIV). With the isolation of a related

5 AIDS-causing virus, the strains originally called HIV are now termed HIV-1 and the related virus is called HIV-2. See, e.g., Guyader et al. (1987) *Nature* 326:662-669; Brun-Vezinet et al. (1986) *Science* 233:343-346; Clavel et al. (1986) *Nature* 324:691-695.

10 A great deal of information has been gathered about the HIV virus, however, to date an effective vaccine has not been identified. Several targets for vaccine development have been examined including the env, Gag, pol and tat gene products encoded by HIV.

15 Haas, et al., (*Current Biology* 6(3):315-324, 1996) suggested that selective codon usage by HIV-1 appeared to account for a substantial fraction of the inefficiency of viral protein synthesis. Andre, et al., (*J. Virol.* 72(2):1497-1503, 1998) described an increased immune

20 response elicited by DNA vaccination employing a synthetic gp120 sequence with optimized codon usage. Schneider, et al., (*J Virol.* 71(7):4892-4903, 1997) discuss inactivation of inhibitory (or instability) elements (INS) located within the coding sequences of the

25 Gag and Gag-protease coding sequences.

The Gag proteins of HIV-1 are necessary for the assembly of virus-like particles. HIV-1 Gag proteins are involved in many stages of the life cycle of the virus including, assembly, virion maturation after particle release, and early post-entry steps in virus replication. The roles of HIV-1 Gag proteins are numerous and complex (Freed, E.O., *Virology* 251:1-15, 1998).

Wolf, et al., (PCT International Application, WO 96/30523, published 3 October 1996; European Patent Application, Publication No. 0 449 116 A1, published 2 October 1991) have described the use of altered pr55 Gag of HIV-1 to act as a non-infectious retroviral-like particulate carrier, in particular, for the presentation of immunologically important epitopes. Wang, et al., (Virology 200:524-534, 1994) describe a system to study assembly of HIV Gag- β -galactosidase fusion proteins into virions. They describe the construction of sequences encoding HIV Gag- β -galactosidase fusion proteins, the expression of such sequences in the presence of HIV Gag proteins, and assembly of these proteins into virus particles.

Recently, Shiver, et al., (PCT International Application, WO 98/34640, published 13 August 1998) described altering HIV-1 (CAM1) Gag coding sequences to produce synthetic DNA molecules encoding HIV Gag and modifications of HIV Gag. The codons of the synthetic molecules were codons preferred by a projected host cell.

The envelope protein of HIV-1 is a glycoprotein of about 160 kD (gp160). During virus infection of the host cell, gp160 is cleaved by host cell proteases to form gp120 and the integral membrane protein, gp41. The gp41 portion is anchored in (and spans) the membrane bilayer of virion, while the gp120 segment protrudes into the surrounding environment. As there is no covalent attachment between gp120 and gp41, free gp120 is released from the surface of virions and infected cells.

Haas, et al., (Current Biology 6(3):315-324, 1996) suggested that selective codon usage by HIV-1 appeared to account for a substantial fraction of the inefficiency of viral protein synthesis. Andre, et al., (J. Virol.

72(2):1497-1503, 1998) described an increased immune response elicited by DNA vaccination employing a synthetic gp120 sequence with optimized codon usage.

5 **SUMMARY OF THE INVENTION**

The present invention relates to improved expression of HIV *Env*-, *tat*-, *pol*-, *prot*-, *reverse transcriptase*, or *Gag*-containing polypeptides and production of virus-like particles.

10 In one embodiment the present invention includes an expression cassette, comprising a polynucleotide encoding an HIV *Gag* polypeptide comprising a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:20. In certain embodiments, the polynucleotide sequence encoding said *Gag* polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:9 or SEQ ID NO:4. The expression cassettes may further include a polynucleotide sequence encoding an HIV *protease* polypeptide, for example a nucleotide sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:5, SEQ ID NO:78, and SEQ ID NO:79. The expression cassettes may further include a polynucleotide sequence encoding an HIV *reverse transcriptase* polypeptide, for example a sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, and SEQ ID NO:84. The expression cassettes may further include a polynucleotide sequence encoding an HIV *tat* polypeptide, for example a sequence selected from the group consisting of: SEQ ID NO:87, SEQ ID NO:88, and SEQ ID NO:89. The expression cassettes may further include a polynucleotide sequence encoding an HIV *polymerase* polypeptide, for example a

sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:6. The expression cassettes may include a polynucleotide sequence encoding an HIV *polymerase* polypeptide, wherein (i) the nucleotide 5 sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:4, and (ii) wherein the sequence is modified by deletions of coding regions corresponding to reverse transcriptase and integrase. The expression 10 cassettes described above may preserves T-helper cell and CTL epitopes. The expression cassettes may further include a polynucleotide sequence encoding an HCV core polypeptide, for example a sequence having at least 90% sequence identity to the sequence presented as SEQ ID 15 NO:7.

In another aspect, the invention includes an expression cassette, comprising a polynucleotide sequence encoding a polypeptide including an HIV *Env* polypeptide, wherein the polynucleotide sequence encoding said *Env* 20 polypeptide comprises a sequence having at least 90% sequence identity to SEQ ID NO:71 (Figure 58) or SEQ ID NO:72 (Figure 59). In certain embodiments, the *Env* expression cassettes includes sequences flanking a V1 region but have a deletion in the V1 region itself, for 25 example the sequence presented as SEQ ID NO:65 (Figure 52, gp160.modUS4.delV1). In certain embodiments, the *Env* expression cassettes, include sequences flanking a V2 region but have a deletion in the V2 region itself, for example the sequences shown in SEQ ID NO:60 (Figure 47); 30 SEQ ID NO:66 (Figure 53); SEQ ID NO:34 (Figure 20); SEQ ID NO:37 (Figure 24); SEQ ID NO:40 (Figure 27); SEQ ID NO:43 (Figure 30); SEQ ID NO:46 (Figure 33); SEQ ID NO:76 (Figure 64) and SEQ ID NO:49 (Figure 36). In certain

embodiments, the Env expression cassettes include sequences flanking a V1/V2 region but have a deletion in the V1/V2 region itself, for example, SEQ ID NO:59 (Figure 46); SEQ ID NO:61 (Figure 48); SEQ ID NO:67 (Figure 54); SEQ ID NO:75 (Figure 63); SEQ ID NO:35 (Figure 21); SEQ ID NO:38 (Figure 25); SEQ ID NO:41 (Figure 28); SEQ ID NO:44 (Figure 31); SEQ ID NO:47 (Figure 34) and SEQ ID NO:50 (Figure 37). The Env-encoding expression cassettes may also include a mutated cleavage site that prevents the cleavage of a gp140 polypeptide into a gp120 polypeptide and a gp41 polypeptide, for example, SEQ ID NO:57 (Figure 44); SEQ ID NO:61 (Figure 48); SEQ ID NO:63 (Figure 50); SEQ ID NO:39 (Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34). The Env expression cassettes may include a gp160 Env polypeptide or a polypeptide derived from a gp160 Env polypeptide, for example SEQ ID NO:64 (Figure 51); SEQ ID NO:65 (Figure 52); SEQ ID NO:66 (Figure 53); SEQ ID NO:67 (Figure 54); SEQ ID NO:68 (Figure 55); SEQ ID NO:75 (Figure 63); SEQ ID NO:73 (Figure 61); SEQ ID NO:48 (Figure 35); SEQ ID NO:49 (Figure 36); SEQ ID NO:50 (Figure 37); SEQ ID NO:76 (Figure 64); and SEQ ID NO:74 (Figure 62). The Env expression cassettes may include a gp140 Env polypeptide or a polypeptide derived from a gp140 Env polypeptide, for example SEQ ID NO:56 (Figure 43); SEQ ID NO:57 (Figure 44); SEQ ID NO:58 (Figure 45); SEQ ID NO:59 (Figure 46); SEQ ID NO:60 (Figure 47); SEQ ID NO:61 (Figure 48); SEQ ID NO:62 (Figure 49); SEQ ID NO:63 (Figure 50); SEQ ID NO:36 (Figure 23); SEQ ID NO:37 (Figure 24); SEQ ID NO:38 (Figure 25); SEQ ID NO:39

(Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34). The Env expression cassettes may also include a gp120 Env polypeptide or a polypeptide derived from a gp120 Env polypeptide, for example SEQ ID NO:54 (Figure 41); and SEQ ID NO:55 (Figure 42); SEQ ID NO:33 (Figure 19); SEQ ID NO:34 (Figure 20); and SEQ ID NO:35 (Figure 21). The Env expression cassettes may include an Env polypeptide lacking the amino acids corresponding to residues 128 to about 194, relative to strains SF162 or US4, for example, SEQ ID NO:55 (Figure 42); SEQ ID NO:62 (Figure 49); SEQ ID NO:63 (Figure 50); and SEQ ID NO:68 (Figure 55).

In another aspect, the invention includes a recombinant expression system for use in a selected host cell, comprising, one or more of the expression cassettes described herein operably linked to control elements compatible with expression in the selected host cell. The expression cassettes may be included on one or on multiple vectors and may use the same or different promoters. Exemplary control elements include a transcription promoter (e.g., CMV, CMV+intron A, SV40, RSV, HIV-Ltr, MMLV-ltr, and metallothionein), a transcription enhancer element, a transcription termination signal, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences.

In another aspect, the invention includes a recombinant expression system for use in a selected host cell, comprising, any one of the expression cassettes described herein operably linked to control elements

compatible with expression in the selected host cell. Exemplary control elements include, but are not limited to, a transcription promoter (e.g., CMV, CMV+intron A, SV40, RSV, HIV-LTR, MMLV-LTR, and metallothionein), a transcription enhancer element, a transcription termination signal, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences.

In yet another aspect, the invention includes a cell comprising one or more of the expression cassettes described herein operably linked to control elements compatible with expression in the cell. The cell can be, for example, a mammalian cell (e.g., BHK, VERO, HT1080, 293, RD, COS-7, or CHO cells), an insect cell (e.g., 15 *Trichoplusia ni* (Tn5) or Sf9), a bacterial cell, a plant cell, a yeast cell, an antigen presenting cell (e.g., primary, immortalized or tumor-derived lymphoid cells such as macrophages, monocytes, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof).

20 In another aspect, the invention includes methods for producing a polypeptide including HIV *Gag*-, *prot*-, *pol*-, reverse transcriptase, *Env*- or *Tat*-containing polypeptide sequences, said method comprising, incubating the cells comprising one or more of the expression cassettes 25 described herein, under conditions for producing said polypeptide.

In yet another aspect, the invention includes compositions for generating an immunological response, comprising one or more of the expression cassettes 30 described herein. In certain embodiments, the compositions also include an adjuvant.

In a still further aspect, the invention includes methods of generating an immune response in a subject, comprising introducing a composition comprising one or

more of the expression cassettes described herein into the subject under conditions that are compatible with expression of said expression cassette in the subject. In certain embodiments, the expression cassette is 5 introduced using a gene delivery vector. More than one expression cassette may be introduced using one or more gene delivery vectors.

In yet another aspect, the invention includes a purified polynucleotide comprising a polynucleotide 10 sequence encoding a polypeptide including an HIV Env polypeptide, wherein the polynucleotide sequence encoding said Env polypeptide comprises a sequence having at least 90% sequence identity to SEQ ID NO:71 (Figure 58) or SEQ ID NO:72 (Figure 59). Further exemplary purified 15 polynucleotide sequences were presented above.

The polynucleotides of the present invention can be produced by recombinant techniques, synthetic techniques, or combinations thereof.

In another embodiment, the invention includes a 20 method for producing a polypeptide including HIV Gag polypeptide sequences, where the method comprises incubating any of the above cells containing an expression cassette of interest under conditions for producing the polypeptide.

25 The invention further includes, a method for producing virus-like particles (VLPs) where the method comprises incubating any of the above-described cells containing an expression cassette of interest under conditions for producing VLPs.

30 In another aspect the invention includes a method for producing a composition of virus-like particles (VLPs) where, any of the above-described cells containing an expression cassette of interest are incubated under

conditions for producing VLPs, and the VLPs are substantially purified to produce a composition of VLPs.

In a further embodiment of the present invention, packaging cell lines are produced using the expression 5 cassettes of the present invention. For example, a cell line useful for packaging lentivirus vectors comprises suitable host cells that have an expression vector containing an expression cassette of the present invention wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the host cell. In a preferred embodiment, such host cells may be transfected with one or more expression cassettes having a polynucleotide sequence that encodes an HIV polymerase polypeptide or 10 polypeptides derived therefrom, for example, where the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:6. Further, the HIV polymerase polypeptide may be modified by deletions of 15 coding regions corresponding to reverse transcriptase and integrase. Such a polynucleotide sequence may preserve T-helper cell and CTL epitopes, for example when used in a vaccine application. In addition, the polynucleotide sequence may also include other polypeptides. Further, 20 polynucleotide sequences encoding additional polypeptides whose expression are useful for packaging cell line function may also be utilized.

In another aspect, the present invention includes a gene delivery or vaccine vector for use in a subject, 30 where the vector is a suitable gene delivery vector for use in the subject, and the vector comprises one or more of any of the expression cassettes of the present

invention where the polynucleotide sequences of interest are operably linked to control elements compatible with expression in the subject. Such gene delivery vectors can be used in a method of DNA immunization of a subject,

5 for example, by introducing a gene delivery vector into the subject under conditions that are compatible with expression of the expression cassette in the subject.

Gene delivery vectors useful in the practice of the present invention include, but are not limited to,

10 nonviral vectors, bacterial plasmid vectors, viral vectors, particulate carriers (where the vector is coated on a polylactide co-glycolide particles, gold or tungsten particle, for example, the coated particle can be delivered to a subject cell using a gene gun), liposome preparations, and viral vectors (e.g., vectors derived from alphaviruses, pox viruses, and vaccinia viruses, as well as, retroviral vectors, including, but not limited to, lentiviral vectors). Alphavirus-derived vectors include, for example, an alphavirus cDNA construct, a

15 recombinant alphavirus particle preparation and a eukaryotic layered vector initiation system. In one embodiment, the subject is a vertebrate, preferably a mammal, and in a further embodiment the subject is a human.

20

25 The invention further includes a method of generating an immune response in a subject, where cells of a subject are transfected with any of the above-described gene delivery vectors (e.g., alphavirus constructs; alphavirus cDNA constructs; eukaryotic layered vector initiation systems (see, e.g., U.S. Patent Number 5,814,482 for description of suitable eukaryotic layered vector initiation systems); alphavirus particle

preparations; etc.) under conditions that permit the expression of a selected polynucleotide and production of a polypeptide of interest (i.e., encoded by any expression cassette of the present invention), thereby eliciting an immunological response to the polypeptide. Transfection of the cells may be performed *ex vivo* and the transfected cells are reintroduced into the subject. Alternately, or in addition, the cells may be transfected *in vivo* in the subject. The immune response may be humoral and/or cell-mediated (cellular).

Further embodiments of the present invention include purified polynucleotides. In one embodiment, the purified polynucleotide comprises a polynucleotide sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:20, and complements thereof. In another embodiment, the purified polynucleotide comprises a polynucleotide sequence encoding an HIV Gag polypeptide, wherein the polynucleotide sequence comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:20, and complements thereof. In still another embodiment, the purified polynucleotide comprises a polynucleotide sequence encoding an HIV Gag polypeptide, wherein the polynucleotide sequence comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:9, and complements thereof. In further embodiments the polynucleotide sequence comprises a sequence having at least 90% sequence identity to one of the following sequences: SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, and complements thereof.

The polynucleotides of the present invention can be produced by recombinant techniques, synthetic techniques, or combinations thereof.

These and other embodiments of the present invention will readily occur to those of ordinary skill in the art in view of the disclosure herein.

5 **BRIEF DESCRIPTION OF THE FIGURES**

Figure 1 shows the locations of the inactivation sites for the native HIV-1SF2 Gag protein coding sequence.

10 Figure 2 shows the locations of the inactivation sites for the native HIV-1SF2 Gag-protease protein coding sequence.

15 Figures 3A and 3B show electron micrographs of virus-like particles. Figure 3A shows immature p55Gag virus-like particles in COS-7 cells transfected with a synthetic HIV-1_{SF2} gag construct while Figure 3B shows mature (arrows) and immature VLP in cells transfected with a modified HIV-1_{SF2} gagprotease construct (GP2, SEQ ID NO:70). Transfected cells were fixed at 24 h (gag) or 48 h (gagprotease) post-transfection and subsequently 20 analyzed by electron microscopy (magnification at 100,000X). Cells transfected with vector alone (pCMVKm2) served as negative control (data not shown).

25 Figure 4 presents an image of samples from a series of fractions which were electrophoresed on an 8-16% SDS polyacrylamide gel and the resulting bands visualized by commassie blue staining. The results show that the native p55 Gag virus-like particles (VLPs) banded at a sucrose density of range of 1.15 - 1.19 g/ml with the peak at approximately 1.17 g/ml.

30 Figure 5 presents an image similar to Figure 4 where the analysis was performed using Gag VLPs produced by a synthetic Gag expression cassette.

Figure 6 presents a comparison of the total amount of purified HIV p55 Gag from several preparations obtained from two baculovirus expression cassettes encoding native and modified Gag.

5 Figure 7 presents an alignment of modified coding sequences of the present invention including a synthetic Gag expression cassette (SEQ ID NO:4), a synthetic Gag-protease expression cassette (SEQ ID NO:5), and a synthetic Gag-polymerase expression cassette (SEQ ID NO:6). A common region (Gag-common; SEQ ID NO:9) extends 10 from position 1 to position 1262.

15 Figure 8 presents an image of wild-type Gag-HCV core expression samples from a series of fractions which were electrophoresed on an 8-16% SDS polyacrylamide gel and the resulting bands visualized by commassie staining.

Figure 9 shows the results of Western blot analysis of the gel shown presented in Figure 8.

20 Figure 10 presents results similar to those shown in Figure 9. The results in Figure 10 indicate that the main HCV Core-specific reactivity migrates at an approximate molecular weight of 72,000 kD, which is in accordance with the predicted molecular weight of the Gag-HCV core chimeric protein.

25 Figures 11A to 11D present a comparison of AT content, in percent, of cDNAs corresponding to an unstable human mRNA (human IFN γ mRNA; 11A), wild-type HIV Gag native RNA (11B), a stable human mRNA (human GAPDH mRNA; 11C), and synthetic HIV Gag RNA (11D).

30 Figure 12 shows the location of the inactivation sites for the native HIV-1SF2 Gag-polymerase sequence.

Figure 13A presents a vector map of pESN2dhfr.

Figure 13B presents a map of the pCMVIII vector.

Figure 14 presents a vector map of pCMV-LINK.

Figure 15 presents a schematic diagram showing the relationships between the following forms of the HIV Env polypeptide: gp160, gp140, gp120, and gp41.

5 Figure 16 depicts the nucleotide sequence of wild-type gp120 from SF162 (SEQ ID NO:30).

Figure 17 depicts the nucleotide sequence of the wild-type gp140 from SF162 (SEQ ID NO:31).

Figure 18 depicts the nucleotide sequence of the wild-type gp160 from SF162 (SEQ ID NO:32).

10 Figure 19 depicts the nucleotide sequence of the construct designated gp120.modSF162 (SEQ ID NO:33).

Figure 20 depicts the nucleotide sequence of the construct designated gp120.modSF162.delV2 (SEQ ID NO:34).

15 Figure 21 depicts the nucleotide sequence of the construct designated gp120.modSF162.delV1/V2 (SEQ ID NO:35).

Figures 22A-H show the percent A-T content over the length of the sequences for IFNY (Figures 2C and 2G); native gp160 Env US4 and SF162 (Figures 2A and 2E, 20 respectively); GAPDH (Figures 2D and 2H); and the synthetic gp160 Env for US4 and SF162 (Figures 2B and 2F, respectively).

Figure 23 depicts the nucleotide sequence of the construct designated gp140.modSF162 (SEQ ID NO:36).

25 Figure 24 depicts the nucleotide sequence of the construct designated gp140.modSF162.delV2 (SEQ ID NO:37).

Figure 25 depicts the nucleotide sequence of the construct designated gp140.modSF162.delV1/V2 (SEQ ID NO:38).

30 Figure 26 depicts the nucleotide sequence of the construct designated gp140.mut.modSF162 (SEQ ID NO:39).

Figure 27 depicts the nucleotide sequence of the construct designated gp140.mut.modSF162.delV2 (SEQ ID NO:40).

Figure 28 depicts the nucleotide sequence of the construct designated gp140.mut.modSF162.delV1/V2 (SEQ ID NO:41).

5 Figure 29 depicts the nucleotide sequence of the construct designated gp140.mut7.modSF162 (SEQ ID NO:42).

Figure 30 depicts the nucleotide sequence of the construct designated gp140.mut7.modSF162.delV2 (SEQ ID NO:43).

10 Figure 31 depicts the nucleotide sequence of the construct designated gp140.mut7.modSF162.delV1/V2 (SEQ ID NO:44).

Figure 32 depicts the nucleotide sequence of the construct designated gp140.mut8.modSF162 (SEQ ID NO:45).

15 Figure 33 depicts the nucleotide sequence of the construct designated gp140.mut8.modSF162.delV2 (SEQ ID NO:46).

Figure 34 depicts the nucleotide sequence of the construct designated gp140.mut8.modSF162.delV1/V2 (SEQ ID NO:47).

20 Figure 35 depicts the nucleotide sequence of the construct designated gp160.modSF162 (SEQ ID NO:48).

Figure 36 depicts the nucleotide sequence of the construct designated gp160.modSF162.delV2 (SEQ ID NO:49).

25 Figure 37 depicts the nucleotide sequence of the construct designated gp160.modSF162.delV1/V2 (SEQ ID NO:50).

Figure 38 depicts the nucleotide sequence of the wild-type gp120 from US4 (SEQ ID NO:51).

30 Figure 39 depicts the nucleotide sequence of the wild-type gp140 from US4 (SEQ ID NO:52).

Figure 40 depicts the nucleotide sequence of the wild-type gp160 from US4 (SEQ ID NO:53).

Figure 41 depicts the nucleotide sequence of the construct designated gp120.modUS4 (SEQ ID NO:54).

Figure 42 depicts the nucleotide sequence of the construct designated gp120.modUS4.del 128-194 (SEQ ID NO:55).

5 Figure 43 depicts the nucleotide sequence of the construct designated gp140.modUS4 (SEQ ID NO:56).

Figure 44 depicts the nucleotide sequence of the construct designated gp140.mut.modUS4 (SEQ ID NO:57).

Figure 45 depicts the nucleotide sequence of the construct designated gp140.TM.modUS4 (SEQ ID NO:58).

10 Figure 46 depicts the nucleotide sequence of the construct designated gp140.modUS4.delV1/V2 (SEQ ID NO:59).

Figure 47 depicts the nucleotide sequence of the construct designated gp140.modUS4.delV2 (SEQ ID NO:60).

15 Figure 48 depicts the nucleotide sequence of the construct designated gp140.mut.modUS4.delV1/V2 (SEQ ID NO:61).

Figure 49 depicts the nucleotide sequence of the construct designated gp140.modUS4.del 128-194 (SEQ ID NO:62).

Figure 50 depicts the nucleotide sequence of the construct designated gp140.mut.modUS4.del 128-194 (SEQ ID NO:63).

25 Figure 51 depicts the nucleotide sequence of the construct designated gp160.modUS4 (SEQ ID NO:64).

Figure 52 depicts the nucleotide sequence of the construct designated gp160.modUS4.delV1 (SEQ ID NO:65).

Figure 53 depicts the nucleotide sequence of the construct designated gp160.modUS4.delV2 (SEQ ID NO:66).

30 Figure 54 depicts the nucleotide sequence of the construct designated gp160.modUS4.delV1/V2 (SEQ ID NO:67).

Figure 55 depicts the nucleotide sequence of the construct designated gp160.modUS4.del 128-194 (SEQ ID NO:68).

5 Figure 56 depicts the nucleotide sequence of the common region of Env from wild-type US4 (SEQ ID NO:69).

Figure 57 depicts the nucleotide sequence of the common region of Env from wild-type SF162 (SEQ ID NO:70).

10 Figure 58 depicts the nucleotide sequence of synthetic sequences corresponding to the common region of Env from US4 (SEQ ID NO:71).

Figure 59 depicts the nucleotide sequence of synthetic sequences corresponding to the common region of Env from SF162 (SEQ ID NO:72).

15 Figure 60 presents a schematic representation of an Env polypeptide purification strategy.

Figure 61 depicts the nucleotide sequence of the bicistronic construct designated gp160.modUS4.Gag.modSF2 (SEQ ID NO:73).

20 Figure 62 depicts the nucleotide sequence of the bicistronic construct designated gp160.modSF162.Gag.modSF2 (SEQ ID NO:74).

Figure 63 depicts the nucleotide sequence of the bicistronic construct designated gp160.modUS4.-delV1/V2.Gag.modSF2 (SEQ ID NO:75).

25 Figure 64 depicts the nucleotide sequence of the bicistronic construct designated gp160.modSF162.delV2.Gag.modSF2 (SEQ ID NO:76).

Figures 65A-65F show micrographs of 293T cells transfected with the following polypeptide encoding 30 sequences: Figure 65A, gag.modSF2; Figure 65B, gp160.modUS4; Figure 65C, gp160.modUS4.delV1/V2.gag.modSF2 (bicistronic Env and Gag); Figures 65D and 65E, gp160.modUS4.delV1/V2 and

gag.modSF2; and Figure 65F, gp120.modSF162.delV2 and gag.modSF2.

Figures 66A and 66B present alignments of selected modified coding sequences of the present invention including a common region defined for each group of synthetic Env expression cassettes. Figure 66A presents alignments of modified SF162 sequences. Figure 66B presents alignments of modified US4 sequences. The SEQ ID NOs for these sequences are presented in Tables 1A and 1B.

Figure 67 shows the ELISA titers (binding antibodies) obtained in two rhesus macaques (H445, lines with solid black dots; and J408, lines with open squares). The y-axis is the end-point gp140 ELISA titers and the x-axis shows weeks post-immunization. The dashed lines at 0, 4, and 8 weeks represent DNA immunizations. The alternating dash/dotted line at 27 weeks indicates a DNA plus protein boost immunization.

Figure 68 (SEQ ID NO:77) depicts the wild-type nucleotide sequence of Gag reverse transcriptase from SF2.

Figure 69 (SEQ ID NO:78) depicts the nucleotide sequence of the construct designated GP1.

Figure 70 (SEQ ID NO:79) depicts the nucleotide sequence of the construct designated GP2.

Figure 71 (SEQ ID NO:80) depicts the nucleotide sequence of the construct designated FS(+).protinact.RTopt.YM. FS(+) indicates that there is a frameshift in the GagPol coding sequence.

Figure 72 (SEQ ID NO:81) depicts the nucleotide sequence of the construct designated FS(+).protinact.RTopt.YMWM.

Figure 73 (SEQ ID NO:82) depicts the nucleotide sequence of the construct designated FS(-

).protmod.RTopt.YM. FS(-) indicates that there is no frameshift in the GagPol coding sequence.

Figure 74 (SEQ ID NO:83) depicts the nucleotide sequence of the construct designated
5 FS(-).protmod.RTopt.YMWM.

Figure 75 (SEQ ID NO:84) depicts the nucleotide sequence of the construct designated FS(-)
).protmod.RTopt(+).

10 Figure 76 (SEQ ID NO:85) depicts the nucleotide sequence of wild type Tat from isolate SF162.

Figure 77 (SEQ ID NO:86) depicts the amino acid sequence of the tat polypeptide.

15 Figure 78 (SEQ ID NO:87) depicts the nucleotide sequence of a synthetic Tat construct designated Tat.SF162.opt.

Figure 79 (SEQ ID NO:88) depicts the nucleotide sequence of a synthetic Tat construct designated tat.cys22.sf162.opt. The construct encodes a tat polypeptide in which the cysteine residue at position 22
20 of the wild type Tat polypeptide is replaced by a glycine residue.

Figures 80A to 80E are an alignment of the nucleotide sequences of the constructs designated Gag.mod.SF2, GP1 (SEQ ID NO:78), and GP2 (SEQ ID NO:79).

25 Figure 81 (SEQ ID NO:89) depicts the nucleotide sequence of the construct designated tataminoSF162.opt, which encodes the amino terminus of that tat protein. The codon encoding the cysteine-22 residue is underlined.

30 Figure 82 (SEQ ID NO:90) depicts the amino acid sequence of the polypeptide encoded by the construct designated tat.cys22.SF162.opt (SEQ ID NO:88).

DETAILED DESCRIPTION OF THE INVENTION

The practice of the present invention will employ, unless otherwise indicated, conventional methods of chemistry, biochemistry, molecular biology, immunology and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); *Methods In Enzymology* (S. Colowick and N. Kaplan, eds., Academic Press, Inc.); and *Handbook of Experimental Immunology*, Vols. I-IV (D.M. Weir and C.C. Blackwell, eds., 1986, Blackwell Scientific Publications); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Short Protocols in Molecular Biology*, 4th ed. (Ausubel et al. eds., 1999, John Wiley & Sons); *Molecular Biology Techniques: An Intensive Laboratory Course*, (Ream et al., eds., 1998, Academic Press); *PCR (Introduction to Biotechniques Series)*, 2nd ed. (Newton & Graham eds., 1997, Springer Verlag).

As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise. Thus, for example, reference to "an antigen" includes a mixture of two or more such agents.

25

1. DEFINITIONS

In describing the present invention, the following terms will be employed, and are intended to be defined as indicated below.

"Synthetic" sequences, as used herein, refers to Env-, tat- or Gag-encoding polynucleotides whose expression has been optimized as described herein, for example, by codon substitution, deletions, replacements and/or inactivation of inhibitory sequences. "Wild-type"

or "native" sequences, as used herein, refers to polypeptide encoding sequences that are essentially as they are found in nature, e.g., Gag encoding sequences as found in the isolate HIV-1SF2 or Env encoding sequences 5 as found in the isolates HIV-1SF162 or HIV1US4.

As used herein, the term "virus-like particle" or "VLP" refers to a nonreplicating, viral shell, derived from any of several viruses discussed further below. VLPs are generally composed of one or more viral 10 proteins, such as, but not limited to those proteins referred to as capsid, coat, shell, surface and/or envelope proteins, or particle-forming polypeptides derived from these proteins. VLPs can form spontaneously upon recombinant expression of the protein in an 15 appropriate expression system. Methods for producing particular VLPs are known in the art and discussed more fully below. The presence of VLPs following recombinant expression of viral proteins can be detected using conventional techniques known in the art, such as by 20 electron microscopy, biophysical characterization, and the like. See, e.g., Baker et al., *Biophys. J.* (1991) 60:1445-1456; Hagensee et al., *J. Virol.* (1994) 68:4503-4505. For example, VLPs can be isolated by density gradient centrifugation and/or identified by 25 characteristic density banding (e.g., Example 7). Alternatively, cryoelectron microscopy can be performed on vitrified aqueous samples of the VLP preparation in question, and images recorded under appropriate exposure conditions.

30 By "particle-forming polypeptide" derived from a particular viral protein is meant a full-length or near full-length viral protein, as well as a fragment thereof, or a viral protein with internal deletions, which has the ability to form VLPs under conditions that favor VLP

formation. Accordingly, the polypeptide may comprise the full-length sequence, fragments, truncated and partial sequences, as well as analogs and precursor forms of the reference molecule. The term therefore intends

5 deletions, additions and substitutions to the sequence, so long as the polypeptide retains the ability to form a VLP. Thus, the term includes natural variations of the specified polypeptide since variations in coat proteins often occur between viral isolates. The term also

10 includes deletions, additions and substitutions that do not naturally occur in the reference protein, so long as the protein retains the ability to form a VLP. Preferred substitutions are those which are conservative in nature, i.e., those substitutions that take place within a family

15 of amino acids that are related in their side chains. Specifically, amino acids are generally divided into four families: (1) acidic -- aspartate and glutamate; (2) basic -- lysine, arginine, histidine; (3) non-polar -- alanine, valine, leucine, isoleucine, proline,

20 phenylalanine, methionine, tryptophan; and (4) uncharged polar -- glycine, asparagine, glutamine, cystine, serine threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified as aromatic amino acids.

25 An "antigen" refers to a molecule containing one or more epitopes (either linear, conformational or both) that will stimulate a host's immune system to make a humoral and/or cellular antigen-specific response. The term is used interchangeably with the term "immunogen."

30 Normally, a B-cell epitope will include at least about 5 amino acids but can be as small as 3-4 amino acids. A T-cell epitope, such as a CTL epitope, will include at least about 7-9 amino acids, and a helper T-cell epitope at least about 12-20 amino acids. Normally, an epitope

will include between about 7 and 15 amino acids, such as, 9, 10, 12 or 15 amino acids. The term "antigen" denotes both subunit antigens, (i.e., antigens which are separate and discrete from a whole organism with which the antigen 5 is associated in nature), as well as, killed, attenuated or inactivated bacteria, viruses, fungi, parasites or other microbes. Antibodies such as anti-idiotype antibodies, or fragments thereof, and synthetic peptide mimotopes, which can mimic an antigen or antigenic 10 determinant, are also captured under the definition of antigen as used herein. Similarly, an oligonucleotide or polynucleotide which expresses an antigen or antigenic determinant *in vivo*, such as in gene therapy and DNA immunization applications, is also included in the 15 definition of antigen herein.

For purposes of the present invention, antigens can be derived from any of several known viruses, bacteria, parasites and fungi, as described more fully below. The term also intends any of the various tumor antigens. 20 Furthermore, for purposes of the present invention, an "antigen" refers to a protein which includes modifications, such as deletions, additions and substitutions (generally conservative in nature), to the native sequence, so long as the protein maintains the 25 ability to elicit an immunological response, as defined herein. These modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through mutations of hosts which produce the antigens.

30 An "immunological response" to an antigen or composition is the development in a subject of a humoral and/or a cellular immune response to an antigen present in the composition of interest. For purposes of the present invention, a "humoral immune response" refers to

an immune response mediated by antibody molecules, while a "cellular immune response" is one mediated by T-lymphocytes and/or other white blood cells. One important aspect of cellular immunity involves an 5 antigen-specific response by cytolytic T-cells ("CTL"s). CTLs have specificity for peptide antigens that are presented in association with proteins encoded by the major histocompatibility complex (MHC) and expressed on the surfaces of cells. CTLs help induce and promote the 10 destruction of intracellular microbes, or the lysis of cells infected with such microbes. Another aspect of cellular immunity involves an antigen-specific response by helper T-cells. Helper T-cells act to help stimulate the function, and focus the activity of, nonspecific 15 effector cells against cells displaying peptide antigens in association with MHC molecules on their surface. A "cellular immune response" also refers to the production of cytokines, chemokines and other such molecules produced by activated T-cells and/or other white blood 20 cells, including those derived from CD4+ and CD8+ T-cells.

A composition or vaccine that elicits a cellular immune response may serve to sensitize a vertebrate subject by the presentation of antigen in association 25 with MHC molecules at the cell surface. The cell-mediated immune response is directed at, or near, cells presenting antigen at their surface. In addition, antigen-specific T-lymphocytes can be generated to allow for the future protection of an immunized host.

The ability of a particular antigen to stimulate a cell-mediated immunological response may be determined by 30 a number of assays, such as by lymphoproliferation (lymphocyte activation) assays, CTL cytotoxic cell assays, or by assaying for T-lymphocytes specific for the

antigen in a sensitized subject. Such assays are well known in the art. See, e.g., Erickson et al., *J. Immunol.* (1993) 151:4189-4199; Doe et al., *Eur. J. Immunol.* (1994) 24:2369-2376. Recent methods of measuring cell-mediated immune response include measurement of intracellular cytokines or cytokine secretion by T-cell populations, or by measurement of epitope specific T-cells (e.g., by the tetramer technique) (reviewed by McMichael, A.J., and O'Callaghan, C.A., *J. Exp. Med.* 187(9):1367-1371, 1998; McHeyzer-Williams, M.G., et al, *Immunol. Rev.* 150:5-21, 1996; Lalvani, A., et al, *J. Exp. Med.* 186:859-865, 1997).

Thus, an immunological response as used herein may be one which stimulates the production of CTLs, and/or the production or activation of helper T-cells. The antigen of interest may also elicit an antibody-mediated immune response. Hence, an immunological response may include one or more of the following effects: the production of antibodies by B-cells; and/or the activation of suppressor T-cells and/or $\gamma\delta$ T-cells directed specifically to an antigen or antigens present in the composition or vaccine of interest. These responses may serve to neutralize infectivity, and/or mediate antibody-complement, or antibody dependent cell cytotoxicity (ADCC) to provide protection to an immunized host. Such responses can be determined using standard immunoassays and neutralization assays, well known in the art.

An "immunogenic composition" is a composition that comprises an antigenic molecule where administration of the composition to a subject results in the development in the subject of a humoral and/or a cellular immune response to the antigenic molecule of interest.

By "subunit vaccine" is meant a vaccine composition which includes one or more selected antigens but not all antigens, derived from or homologous to, an antigen from a pathogen of interest such as from a virus, bacterium, parasite or fungus. Such a composition is substantially free of intact pathogen cells or pathogenic particles, or the lysate of such cells or particles. Thus, a "subunit vaccine" can be prepared from at least partially purified (preferably substantially purified) immunogenic polypeptides from the pathogen, or analogs thereof. The method of obtaining an antigen included in the subunit vaccine can thus include standard purification techniques, recombinant production, or synthetic production.

"Substantially purified" general refers to isolation of a substance (compound, polynucleotide, protein, polypeptide, polypeptide composition) such that the substance comprises the majority percent of the sample in which it resides. Typically in a sample a substantially purified component comprises 50%, preferably 80%-85%, more preferably 90-95% of the sample. Techniques for purifying polynucleotides and polypeptides of interest are well-known in the art and include, for example, ion-exchange chromatography, affinity chromatography and sedimentation according to density.

A "coding sequence" or a sequence which "encodes" a selected polypeptide, is a nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences (or "control elements"). The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but

is not limited to, cDNA from viral, procaryotic or eucaryotic mRNA, genomic DNA sequences from viral or procaryotic DNA, and even synthetic DNA sequences. A transcription termination sequence may be located 3' to 5 the coding sequence.

Typical "control elements", include, but are not limited to, transcription promoters, transcription enhancer elements, transcription termination signals, polyadenylation sequences (located 3' to the translation 10 stop codon), sequences for optimization of initiation of translation (located 5' to the coding sequence), and translation termination sequences, see e.g., McCaughey et al. (1995) *PNAS USA* 92:5431-5435; Kochetov et al (1998) *FEBS Letts.* 440:351-355.

15 A "nucleic acid" molecule can include, but is not limited to, procaryotic sequences, eucaryotic mRNA, cDNA from eucaryotic mRNA, genomic DNA sequences from eucaryotic (e.g., mammalian) DNA, and even synthetic DNA sequences. The term also captures sequences that include 20 any of the known base analogs of DNA and RNA.

"Operably linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, a given promoter operably linked to a coding sequence is 25 capable of effecting the expression of the coding sequence when the proper enzymes are present. The promoter need not be contiguous with the coding sequence, so long as it functions to direct the expression thereof. Thus, for example, intervening untranslated yet 30 transcribed sequences can be present between the promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

"Recombinant" as used herein to describe a nucleic acid molecule means a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of the polynucleotide with which it is associated in nature; and/or (2) is linked to a polynucleotide other than that to which it is linked in nature. The term "recombinant" as used with respect to a protein or polypeptide means a polypeptide produced by expression of a recombinant polynucleotide. "Recombinant host cells," "host cells," "cells," "cell lines," "cell cultures," and other such terms denoting prokaryotic microorganisms or eukaryotic cell lines cultured as unicellular entities, are used interchangeably, and refer to cells which can be, or have been, used as recipients for recombinant vectors or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to accidental or deliberate mutation. Progeny of the parental cell which are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a desired peptide, are included in the progeny intended by this definition, and are covered by the above terms.

Techniques for determining amino acid sequence "similarity" are well known in the art. In general, "similarity" means the exact amino acid to amino acid comparison of two or more polypeptides at the appropriate place, where amino acids are identical or possess similar chemical and/or physical properties such as charge or hydrophobicity. A so-termed "percent similarity" then

can be determined between the compared polypeptide sequences. Techniques for determining nucleic acid and amino acid sequence identity also are well known in the art and include determining the nucleotide sequence of 5 the mRNA for that gene (usually via a cDNA intermediate) and determining the amino acid sequence encoded thereby, and comparing this to a second amino acid sequence. In general, "identity" refers to an exact nucleotide to nucleotide or amino acid to amino acid correspondence of 10 two polynucleotides or polypeptide sequences, respectively.

Two or more polynucleotide sequences can be compared by determining their "percent identity." Two or more amino acid sequences likewise can be compared by 15 determining their "percent identity." The percent identity of two sequences, whether nucleic acid or peptide sequences, is generally described as the number of exact matches between two aligned sequences divided by the length of the shorter sequence and multiplied by 100. 20 An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981). This algorithm can be extended to use with peptide sequences using the scoring matrix developed by 25 Dayhoff, Atlas of Protein Sequences and Structure, M.O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, Nucl. Acids Res. 14(6):6745-6763 (1986). An implementation of this algorithm for nucleic 30 acid and peptide sequences is provided by the Genetics Computer Group (Madison, WI) in their BestFit utility application. The default parameters for this method are

described in the Wisconsin Sequence Analysis Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, WI). Other equally suitable programs for calculating the percent identity or 5 similarity between sequences are generally known in the art.

For example, percent identity of a particular nucleotide sequence to a reference sequence can be determined using the homology algorithm of Smith and Waterman with a default scoring table and a gap penalty of six nucleotide positions. Another method of establishing percent identity in the context of the present invention is to use the MPSRCH package of programs copyrighted by the University of Edinburgh, 10 developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages, the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap 15 extension penalty of one, and a gap of six). From the data generated, the "Match" value reflects "sequence identity." Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, such as the alignment program BLAST, which can also be used with default parameters. 20 For example, BLASTN and BLASTP can be used with the following default parameters: genetic code = standard; filter = none; strand = both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by = 25 HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details of these programs can be found at 30

the following internet address:

<http://www.ncbi.nlm.gov/cgi-bin/BLAST>.

One of skill in the art can readily determine the proper search parameters to use for a given sequence in
5 the above programs. For example, the search parameters may vary based on the size of the sequence in question. Thus, for example, a representative embodiment of the present invention would include an isolated polynucleotide having X contiguous nucleotides, wherein
10 (i) the X contiguous nucleotides have at least about 50% identity to Y contiguous nucleotides derived from any of the sequences described herein, (ii) X equals Y, and (iii) X is greater than or equal to 6 nucleotides and up to 5000 nucleotides, preferably greater than or equal to
15 8 nucleotides and up to 5000 nucleotides, more preferably 10-12 nucleotides and up to 5000 nucleotides, and even more preferably 15-20 nucleotides, up to the number of nucleotides present in the full-length sequences described herein (e.g., see the Sequence Listing and
20 claims), including all integer values falling within the above-described ranges.

The synthetic expression cassettes (and purified polynucleotides) of the present invention include related polynucleotide sequences having about 80% to 100%,
25 greater than 80-85%, preferably greater than 90-92%, more preferably greater than 95%, and most preferably greater than 98% sequence (including all integer values falling within these described ranges) identity to the synthetic expression cassette sequences disclosed herein (for example, to the sequences presented in Tables 1A and 1B) when the sequences of the present invention are used as the query sequence.

Two nucleic acid fragments are considered to "selectively hybridize" as described herein. The degree of sequence identity between two nucleic acid molecules affects the efficiency and strength of hybridization events between such molecules. A partially identical nucleic acid sequence will at least partially inhibit a completely identical sequence from hybridizing to a target molecule. Inhibition of hybridization of the completely identical sequence can be assessed using hybridization assays that are well known in the art (e.g., Southern blot, Northern blot, solution hybridization, or the like, see Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, (1989) Cold Spring Harbor, N.Y.). Such assays can be conducted using varying degrees of selectivity, for example, using conditions varying from low to high stringency. If conditions of low stringency are employed, the absence of non-specific binding can be assessed using a secondary probe that lacks even a partial degree of sequence identity (for example, a probe having less than about 30% sequence identity with the target molecule), such that, in the absence of non-specific binding events, the secondary probe will not hybridize to the target.

When utilizing a hybridization-based detection system, a nucleic acid probe is chosen that is complementary to a target nucleic acid sequence, and then by selection of appropriate conditions the probe and the target sequence "selectively hybridize," or bind, to each other to form a hybrid molecule. A nucleic acid molecule that is capable of hybridizing selectively to a target sequence under "moderately stringent" typically

hybridizes under conditions that allow detection of a target nucleic acid sequence of at least about 10-14 nucleotides in length having at least approximately 70% sequence identity with the sequence of the selected 5 nucleic acid probe. Stringent hybridization conditions typically allow detection of target nucleic acid sequences of at least about 10-14 nucleotides in length having a sequence identity of greater than about 90-95% with the sequence of the selected nucleic acid probe.

10 Hybridization conditions useful for probe/target hybridization where the probe and target have a specific degree of sequence identity, can be determined as is known in the art (see, for example, Nucleic Acid Hybridization: A Practical Approach, editors B.D. Hames 15 and S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

With respect to stringency conditions for hybridization, it is well known in the art that numerous equivalent conditions can be employed to establish a 20 particular stringency by varying, for example, the following factors: the length and nature of probe and target sequences, base composition of the various sequences, concentrations of salts and other hybridization solution components, the presence or 25 absence of blocking agents in the hybridization solutions (e.g., formamide, dextran sulfate, and polyethylene glycol), hybridization reaction temperature and time parameters, as well as, varying wash conditions. The selection of a particular set of hybridization conditions 30 is selected following standard methods in the art (see, for example, Sambrook, et al., Molecular Cloning: A

Laboratory Manual, Second Edition, (1989) Cold Spring Harbor, N.Y.).

A first polynucleotide is "derived from" second polynucleotide if it has the same or substantially the
5 same basepair sequence as a region of the second polynucleotide, its cDNA, complements thereof, or if it displays sequence identity as described above.

A first polypeptide is "derived from" a second polypeptide if it is (i) encoded by a first
10 polynucleotide derived from a second polynucleotide, or (ii) displays sequence identity to the second polypeptides as described above.

Generally, a viral polypeptide is "derived from" a particular polypeptide of a virus (viral polypeptide) if
15 it is (i) encoded by an open reading frame of a polynucleotide of that virus (viral polynucleotide), or (ii) displays sequence identity to polypeptides of that virus as described above.

"Encoded by" refers to a nucleic acid sequence which
20 codes for a polypeptide sequence, wherein the polypeptide sequence or a portion thereof contains an amino acid sequence of at least 3 to 5 amino acids, more preferably at least 8 to 10 amino acids, and even more preferably at least 15 to 20 amino acids from a polypeptide encoded by
25 the nucleic acid sequence. Also encompassed are polypeptide sequences which are immunologically identifiable with a polypeptide encoded by the sequence.

"Purified polynucleotide" refers to a polynucleotide of interest or fragment thereof which is essentially free, e.g., contains less than about 50%, preferably less than about 70%, and more preferably less than about 90%, of the protein with which the polynucleotide is naturally associated. Techniques for purifying polynucleotides of interest are well-known in the art and include, for

example, disruption of the cell containing the polynucleotide with a chaotropic agent and separation of the polynucleotide(s) and proteins by ion-exchange chromatography, affinity chromatography and sedimentation according to density.

By "nucleic acid immunization" is meant the introduction of a nucleic acid molecule encoding one or more selected antigens into a host cell, for the *in vivo* expression of an antigen, antigens, an epitope, or epitopes. The nucleic acid molecule can be introduced directly into a recipient subject, such as by injection, inhalation, oral, intranasal and mucosal administration, or the like, or can be introduced *ex vivo*, into cells which have been removed from the host. In the latter case, the transformed cells are reintroduced into the subject where an immune response can be mounted against the antigen encoded by the nucleic acid molecule.

"Gene transfer" or "gene delivery" refers to methods or systems for reliably inserting DNA or RNA of interest into a host cell. Such methods can result in transient expression of non-integrated transferred DNA, extrachromosomal replication and expression of transferred replicons (e.g., episomes), or integration of transferred genetic material into the genomic DNA of host cells. Gene delivery expression vectors include, but are not limited to, vectors derived from bacterial plasmid vectors, viral vectors, non-viral vectors, alphaviruses, pox viruses and vaccinia viruses. When used for immunization, such gene delivery expression vectors may be referred to as vaccines or vaccine vectors.

"T lymphocytes" or "T cells" are non-antibody producing lymphocytes that constitute a part of the cell-mediated arm of the immune system. T cells arise from immature lymphocytes that migrate from the bone marrow to

the thymus, where they undergo a maturation process under the direction of thymic hormones. Here, the mature lymphocytes rapidly divide increasing to very large numbers. The maturing T cells become immunocompetent based on their ability to recognize and bind a specific antigen. Activation of immunocompetent T cells is triggered when an antigen binds to the lymphocyte's surface receptors.

The term "transfection" is used to refer to the uptake of foreign DNA by a cell. A cell has been "transfected" when exogenous DNA has been introduced inside the cell membrane. A number of transfection techniques are generally known in the art. See, e.g., Graham et al. (1973) *Virology*, 52:456, Sambrook et al. (1989) *Molecular Cloning, a laboratory manual*, Cold Spring Harbor Laboratories, New York, Davis et al. (1986) *Basic Methods in Molecular Biology*, Elsevier, and Chu et al. (1981) *Gene* 13:197. Such techniques can be used to introduce one or more exogenous DNA moieties into suitable host cells. The term refers to both stable and transient uptake of the genetic material, and includes uptake of peptide- or antibody-linked DNAs.

A "vector" is capable of transferring gene sequences to target cells (e.g., bacterial plasmid vectors, viral vectors, non-viral vectors, particulate carriers, and liposomes). Typically, "vector construct," "expression vector," and "gene transfer vector," mean any nucleic acid construct capable of directing the expression of a gene of interest and which can transfer gene sequences to target cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors.

Transfer of a "suicide gene" (e.g., a drug-susceptibility gene) to a target cell renders the cell sensitive to compounds or compositions that are

relatively nontoxic to normal cells. Moolten, F.L. (1994) *Cancer Gene Ther.* 1:279-287. Examples of suicide genes are thymidine kinase of herpes simplex virus (HSV-tk), cytochrome P450 (Manome et al. (1996) *Gene Therapy* 3:513-520), human deoxycytidine kinase (Manome et al. (1996) *Nature Medicine* 2(5):567-573) and the bacterial enzyme cytosine deaminase (Dong et al. (1996) *Human Gene Therapy* 7:713-720). Cells which express these genes are rendered sensitive to the effects of the relatively nontoxic prodrugs ganciclovir (HSV-tk), cyclophosphamide (cytochrome P450 2B1), cytosine arabinoside (human deoxycytidine kinase) or 5-fluorocytosine (bacterial cytosine deaminase). Culver et al. (1992) *Science* 256:1550-1552, Huber et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:8302-8306.

A "selectable marker" or "reporter marker" refers to a nucleotide sequence included in a gene transfer vector that has no therapeutic activity, but rather is included to allow for simpler preparation, manufacturing, characterization or testing of the gene transfer vector.

A "specific binding agent" refers to a member of a specific binding pair of molecules wherein one of the molecules specifically binds to the second molecule through chemical and/or physical means. One example of a specific binding agent is an antibody directed against a selected antigen.

By "subject" is meant any member of the subphylum chordata, including, without limitation, humans and other primates, including non-human primates such as chimpanzees and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats and guinea pigs; birds, including domestic, wild and game birds such

as chickens, turkeys and other gallinaceous birds, ducks, geese, and the like. The term does not denote a particular age. Thus, both adult and newborn individuals, are intended to be covered. The system described above 5 is intended for use in any of the above vertebrate species, since the immune systems of all of these vertebrates operate similarly.

By "pharmaceutically acceptable" or "pharmacologically acceptable" is meant a material which 10 is not biologically or otherwise undesirable, i.e., the material may be administered to an individual in a formulation or composition without causing any undesirable biological effects or interacting in a deleterious manner with any of the components of the 15 composition in which it is contained.

By "physiological pH" or a "pH in the physiological range" is meant a pH in the range of approximately 7.2 to 8.0 inclusive, more typically in the range of approximately 7.2 to 7.6 inclusive.

20 As used herein, "treatment" refers to any of (I) the prevention of infection or reinfection, as in a traditional vaccine, (ii) the reduction or elimination of symptoms, and (iii) the substantial or complete elimination of the pathogen in question. Treatment may 25 be effected prophylactically (prior to infection) or therapeutically (following infection).

"Lentiviral vector", and "recombinant lentiviral vector" are derived from the subset of retroviral vectors known as lentiviruses. Lentiviral vectors refer to a 30 nucleic acid construct which carries, and within certain embodiments, is capable of directing the expression of a nucleic acid molecule of interest. The lentiviral vector includes at least one transcriptional promoter/enhancer or locus defining element(s), or other elements which

control gene expression by other means such as alternate splicing, nuclear RNA export, post-translational modification of messenger, or post-transcriptional modification of protein. Such vector constructs must 5 also include a packaging signal, long terminal repeats (LTRS) or portion thereof, and positive and negative strand primer binding sites appropriate to the lentiviral vector used (if these are not already present in the retroviral vector). Optionally, the recombinant 10 lentiviral vector may also include a signal which directs polyadenylation, selectable markers such as Neo, TK, hygromycin, phleomycin, histidinol, or DHFR, as well as one or more restriction sites and a translation termination sequence. By way of example, such vectors 15 typically include a 5' LTR, a tRNA binding site, a packaging signal, an origin of second strand DNA synthesis, and a 3'LTR or a portion thereof.

"Lentiviral vector particle" as utilized within the present invention refers to a lentivirus which carries at 20 least one gene of interest. The retrovirus may also contain a selectable marker. The recombinant lentivirus is capable of reverse transcribing its genetic material (RNA) into DNA and incorporating this genetic material into a host cell's DNA upon infection. Lentiviral vector 25 particles may have a lentiviral envelope, a non-lentiviral envelope (e.g., an ambo or VSV-G envelope), or a chimeric envelope.

"Nucleic acid expression vector" or "Expression cassette" refers to an assembly which is capable of 30 directing the expression of a sequence or gene of interest. The nucleic acid expression vector includes a promoter which is operably linked to the sequences or gene(s) of interest. Other control elements may be present as well. Expression cassettes described herein

may be contained within a plasmid construct. In addition to the components of the expression cassette, the plasmid construct may also include a bacterial origin of replication, one or more selectable markers, a signal 5 which allows the plasmid construct to exist as single-stranded DNA (e.g., a M13 origin of replication), a multiple cloning site, and a "mammalian" origin of replication (e.g., a SV40 or adenovirus origin of replication).

10 "Packaging cell" refers to a cell which contains those elements necessary for production of infectious recombinant retrovirus (e.g., lentivirus) which are lacking in a recombinant retroviral vector. Typically, such packaging cells contain one or more expression 15 cassettes which are capable of expressing proteins which encode Gag, pol and env proteins.

"Producer cell" or "vector producing cell" refers to a cell which contains all elements necessary for production of recombinant retroviral vector particles.

20

2. MODES OF CARRYING OUT THE INVENTION

Before describing the present invention in detail, it is to be understood that this invention is not limited to particular formulations or process parameters as such 25 may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not intended to be limiting.

Although a number of methods and materials similar 30 or equivalent to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein.

2.1 SYNTHETIC EXPRESSION CASSETTES**2.1.1 MODIFICATION OF HIV-1 GAG NUCLEIC ACID CODING SEQUENCES**

One aspect of the present invention is the
5 generation of HIV-1 Gag protein coding sequences, and
related sequences, having improved expression relative to
the corresponding wild-type sequence. An exemplary
embodiment of the present invention is illustrated herein
modifying the Gag protein wild-type sequences obtained
10 from the HIV-1SF2 strain (SEQ ID NO:1; Sanchez-Pescador,
R., et al., *Science* 227(4686): 484-492, 1985; Luciw,
P.A., et al. U.S. Patent No. 5,156,949, issued October
20, 1992; Luciw, P.A., et al., U.S. Patent No. 5,688,688,
November 18, 1997). Gag sequence obtained from other HIV
15 variants may be manipulated in similar fashion following
the teachings of the present specification. Such other
variants include, but are not limited to, Gag protein
encoding sequences obtained from the isolates HIV_{IIIB},
HIV_{SF2}, HIV-
20 1_{SF162}, HIV-1_{SF170}, HIV_{LAV}, HIV_{LAI}, HIV_{MN}, HIV-1_{CM235}, HIV-1_{US4},
other HIV-1 strains from diverse subtypes (e.g.,
subtypes, A through G, and O), HIV-2 strains and diverse
subtypes (e.g., HIV-2_{UC1} and HIV-2_{UC2}), and simian
immunodeficiency virus (SIV). (See, e.g., *Virology*, 3rd
25 Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd
Edition (B.N. Fields and D.M. Knipe, eds. 1991);
Virology, 3rd Edition (Fields, BN, DM Knipe, PM Howley,
Editors, 1996, Lippincott-Raven, Philadelphia, PA; for a
description of these and other related viruses).
30 First, the HIV-1 codon usage pattern was modified so
that the resulting nucleic acid coding sequence was
comparable to codon usage found in highly expressed human
genes (Example 1). The HIV codon usage reflects a high
content of the nucleotides A or T of the codon-triplet.

The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the 5 nucleotides G or C. The Gag coding sequences were modified to be comparable to codon usage found in highly expressed human genes. In Figure 11 (Example 1), the percent A-T content of cDNA sequences corresponding to the mRNA for a known unstable mRNA and a known stable 10 mRNA are compared to the percent A-T content of native HIV-1SF2 Gag cDNA and to the synthetic Gag cDNA sequence of the present invention. Experiments performed in support of the present invention showed that the synthetic Gag sequences were capable of higher level of 15 protein production (see the Examples) relative to the native Gag sequences. The data in Figure 11 suggest that one reason for this increased production is increased stability of the mRNA corresponding to the synthetic Gag coding sequences versus the mRNA corresponding to the 20 native Gag coding sequences.

Second, there are inhibitory (or instability) elements (INS) located within the coding sequences of the Gag coding sequences (Example 1). The RRE is a secondary RNA structure that interacts with the HIV encoded Rev- 25 protein to overcome the expression down-regulating effects of the INS. To overcome the post-transcriptional activating mechanisms of RRE and Rev, the instability elements were inactivated by introducing multiple point mutations that did not alter the reading frame of the 30 encoded proteins. Figure 1 shows the original SF2 Gag sequence, the location of the INS sequences, and the modifications made to the INS sequences to reduce their effects. The resulting modified coding sequences are

presented as a synthetic Gag expression cassette (SEQ ID NO:4).

Modification of the Gag polypeptide coding sequences resulted in improved expression relative to the wild-type 5 coding sequences in a number of mammalian cell lines (as well as other types of cell lines, including, but not limited to, insect cells). Further, expression of the sequences resulted in production of virus-like particles (VLPs) by these cell lines (see below). Similar Gag 10 polypeptide coding sequences can be obtained from a variety of isolates (families, sub-types, strains, etc.) including, but not limited to such other variants include, but are not limited to, Gag polypeptide encoding 15 sequences obtained from the isolates HIV_{IIIB}, HIV_{SF2}, HIV-1_{SF162}, HIV-1_{SF170}, HIV_{LAV}, HIV_{LAI}, HIV_{MN}, HIV-1_{CM235}, HIV-1_{US4}, other HIV-1 strains from diverse subtypes (e.g., subtypes, A through G, and O), HIV-2 strains and diverse subtypes (e.g., HIV-2_{UC1} and HIV-2_{UC2}), and simian immunodeficiency virus (SIV). (See, e.g., Virology, 3rd Edition (W.K. 20 Joklik ed. 1988); Fundamental Virology, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991; Virology, 3rd Edition (Fields, BN, DM Knipe, PM Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA). Gag polypeptide 25 encoding sequences derived from these variants can be optimized and tested for improved expression in mammals by following the teachings of the present specification (see the Examples, in particular Example 1).

2.1.2 FURTHER MODIFICATION OF SEQUENCES INCLUDING HIV-1
30 GAG NUCLEIC ACID CODING SEQUENCES

Experiments performed in support of the present invention have shown that similar modifications of HIV-1 Gag-protease, Gag-reverse transcriptase and Gag-polymerase sequences also result in improved expression

of the polyproteins, as well as, the production of VLPs formed by polypeptides produced from such modified coding sequences.

For the Gag-protease sequence (wild type, SEQ ID NO:2; modified, SEQ ID NOS:5, 78, 79), the changes in codon usage were restricted to the regions upstream of the -1 frameshift (Figure 2). Further, inhibitory (or instability) elements (INS) located within the coding sequences of the Gag-protease polypeptide coding sequence were altered as well (indicated in Figure 2). Exemplary constructs (which include the -1 frameshift) encoding modified Gag-protease sequences include those shown in SEQ ID NOS:78 and 79 (Figures 69 and 70). These are: GP1 (SEQ ID NO:78) in which the protease region was also codon optimized and INS inactivated and GP2 (SEQ ID NO:79), in which the protease region was only subjected to INS inactivation.

For other Gag-containing sequences, for example the Gag-polymerase sequence (wild type, SEQ ID NO:3; modified, SEQ ID NO:6) or Gag-reverse transcriptase (wild type, SEQ ID NO:77; modified SEQ ID NOS:80-84), the changes in codon usage are similar to those for the Gag-protease sequence. Those expression cassettes which contain a frameshift in the GagPol coding sequence are designated "FS(+)" (SEQ ID NOS:80 and 81, Figures 71 and 72) while the designation "FS(-)" (SEQ ID Nos: 82, 83 and 84, Figures 73, 74 and 75) indicates that there is no frameshift utilized in this coding sequence.

In addition to polyproteins containing HIV-related sequences, the various Gag-, Gag-prot, Gag-pol, Gag-reverse transcriptase encoding sequences of the present invention can be fused to other polypeptides (creating chimeric polypeptides) for which an immunogenic response is desired. An example of such a chimeric protein is the

joining of the improved expression Gag encoding sequences to the Hepatitis C Virus (HCV) core protein. In this case, the HCV-core encoding sequences were placed in-frame with the HIV-Gag encoding sequences, resulting in 5 the Gag/HCV-core encoding sequence presented as SEQ ID NO:7 (wild type sequence presented as SEQ ID NO:8).

Further sequences useful in the practice of the present invention include, but are not limited to, sequences encoding viral epitopes/antigens {including but 10 not limited to, HCV antigens (e.g., E1, E2; Houghton, M., et al., U.S. Patent No. 5,714,596, issued February 3, 1998; Houghton, M., et al., U.S. Patent No. 5,712,088, issued January 27, 1998; Houghton, M., et al., U.S. Patent No. 5,683,864, issued November 4, 1997; 15 Weiner, A.J., et al., U.S. Patent No. 5,728,520, issued March 17, 1998; Weiner, A.J., et al., U.S. Patent No. 5,766,845, issued June 16, 1998; Weiner, A.J., et al., U.S. Patent No. 5,670,152, issued September 23, 1997), HIV antigens (e.g., derived from nef, tat, rev, vpu, vif, 20 vpr and/or env); and sequences encoding tumor antigens/epitopes. Additional sequences are described below. Also, variations on the orientation of the Gag and other coding sequences, relative to each other, are also described below.

25 Gag, Gag-protease, Gag-reverse transcriptase and/or Gag-polymerase polypeptide coding sequences can be obtained from any HIV isolates (different families, subtypes, and strains) including but not limited to the isolates HIV_{IIIB}, HIV_{SF2}, HIV_{SP162}, HIVus4, HIV_{cm235}, HIV_{LAV}, 30 HIV_{LAI}, HIV_{MN} (see, e.g., Myers et al. Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers et al., *Human Retroviruses and Aids*, 1997, Los Alamos, New Mexico: Los Alamos National Laboratory). Synthetic expression cassettes can be generated using

such coding sequences as starting material by following the teachings of the present specification (e.g., see Example 1). Further, the synthetic expression cassettes of the present invention include related Gag polypeptide 5 coding sequences having greater than 75%, preferably greater than 80-85%, more preferably greater than 90-95%, and most preferably greater than 98% sequence identity (or any integer value within these ranges) to the synthetic expression cassette sequences disclosed herein 10 (for example, SEQ ID NO:4; SEQ ID NO:5; SEQ ID NO:6; and SEQ ID NO:20, the Gag Major Homology Region).

2.1.3 EXPRESSION OF SYNTHETIC SEQUENCES ENCODING HIV-1 GAG AND RELATED POLYPEPTIDES

15 Several synthetic Gag-encoding sequences (expression cassettes) of the present invention were cloned into a number of different expression vectors (Example 1) to evaluate levels of expression and production of VLPs. Two modified synthetic coding sequences are presented as 20 a synthetic Gag expression cassette (SEQ ID NO:4) and a synthetic Gag-protease expression cassette (SEQ ID NOS:78 and 79). Other synthetic Gag-encoding proteins are presented, for example, as SEQ ID NOS:80 through 84. The 25 synthetic DNA fragments for Gag-encoding polypeptides (e.g., Gag, Gag-protease, Gag-polymerase, Gag-reverse transcriptase) were cloned into expression vectors described in Example 1, including, a transient expression vector, CMV-promoter-based mammalian vectors, and a 30 shuttle vector for use in baculovirus expression systems. Corresponding wild-type sequences were cloned into the same vectors.

These vectors were then transfected into a several different cell types, including a variety of mammalian

cell lines, (293, RD, COS-7, and CHO, cell lines available, for example, from the A.T.C.C.). The cell lines were cultured under appropriate conditions and the levels of p24 (Gag) expression in supernatants were 5 evaluated (Example 2). The results of these assays demonstrated that expression of synthetic Gag-encoding sequences were significantly higher than corresponding wild-type sequences (Example 2; Table 2).

Further, Western Blot analysis showed that cells 10 containing the synthetic Gag expression cassette produced the expected 55 kD (p55) protein at higher per-cell concentrations than cells containing the native expression cassette. The Gag p55 protein was seen in both cell lysates and supernatants. The levels of 15 production were significantly higher in cell supernatants for cells transfected with the synthetic Gag expression cassette of the present invention. Experiments performed in support of the present invention suggest that cells containing the synthetic Gag-prot expression cassettes 20 produced the expected Gag-prot protein at comparably higher per-cell concentrations than cells containing the wild-type expression cassette.

Fractionation of the supernatants from mammalian 25 cells transfected with the synthetic Gag expression cassette showed that it provides superior production of both p55 protein and VLPs, relative to the wild-type Gag sequences (Examples 6 and 7).

Efficient expression of these Gag-containing polypeptides in mammalian cell lines provides the 30 following benefits: the Gag polypeptides are free of baculovirus contaminants; production by established methods approved by the FDA; increased purity; greater yields (relative to native coding sequences); and a novel method of producing the Gag-containing polypeptides in

CHO or other mammalian cells which is not feasible in the absence of the increased expression obtained using the constructs of the present invention. Exemplary Mammalian cell lines include, but are not limited to, BHK, VERO, 5 HT1080, 293, 293T, RD, COS-7, CHO, Jurkat, HUT, SUPT, C8166, MOLT4/clone8, MT-2, MT-4, H9, PM1, CEM, myeloma cells (e.g., SB20 cells) and CEMX174, such cell lines are available, for example, from the A.T.C.C.).

A synthetic Gag expression cassette of the present 10 invention also demonstrated high levels of expression and VLP production when transfected into insect cells (Example 7). Further, in addition to a higher total protein yield, the final product from the synthetic p55-expressed Gag consistently contained lower amounts of 15 contaminating baculovirus proteins than the final purified product from the native p55-expressed Gag.

Further, synthetic Gag expression cassettes of the present invention have also been introduced into yeast vectors which were transformed into and efficiently 20 expressed by yeast cells (*Saccharomyces cerevisea*; using vectors as described in Rosenberg, S. and Tekamp-Olson, P., U.S. Patent No. RE35,749, issued, March 17, 1998).

In addition to the mammalian and insect vectors 25 described in the Examples, the synthetic expression cassettes of the present invention can be incorporated into a variety of expression vectors using selected expression control elements. Appropriate vectors and control elements for any given cell type can be selected by one having ordinary skill in the art in view of the 30 teachings of the present specification and information known in the art about expression vectors.

For example, a synthetic Gag expression cassette can be inserted into a vector which includes control elements operably linked to the desired coding sequence, which

allow for the expression of the gene in a selected cell-type. For example, typical promoters for mammalian cell expression include the SV40 early promoter, a CMV promoter such as the CMV immediate early promoter (a CMV promoter can include intron A), RSV, HIV-LTR, the mouse mammary tumor virus LTR promoter (MMLV-LTR), FIV-LTR, the adenovirus major late promoter (Ad MLP), and the herpes simplex virus promoter, among others. Other nonviral promoters, such as a promoter derived from the murine metallothionein gene, will also find use for mammalian expression. Typically, transcription termination and polyadenylation sequences will also be present, located 5' to the translation stop codon. Preferably, a sequence for optimization of initiation of translation, located 5' to the coding sequence, is also present. Examples of transcription terminator/polyadenylation signals include those derived from SV40, as described in Sambrook, et al., *supra*, as well as a bovine growth hormone terminator sequence. Introns, containing splice donor and acceptor sites, may also be designed into the constructs for use with the present invention (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986).

Enhancer elements may also be used herein to increase expression levels of the mammalian constructs. Examples include the SV40 early gene enhancer, as described in Dijkema et al., *EMBO J.* (1985) 4:761, the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus, as described in Gorman et al., *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777 and elements derived from human CMV, as described in Boshart et al., *Cell* (1985) 41:521, such as elements included in the CMV intron A sequence (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986).

The desired synthetic Gag polypeptide encoding sequences can be cloned into any number of commercially available vectors to generate expression of the polypeptide in an appropriate host system. These systems 5 include, but are not limited to, the following:

baculovirus expression {Reilly, P.R., et al., BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992); Beames, et al., Biotechniques 11:378 (1991); Pharmingen; Clontech, Palo Alto, CA)}, vaccinia expression {Earl, P. L., et al., "Expression of proteins in mammalian cells using 10 vaccinia" In Current Protocols in Molecular Biology (F. M. Ausubel, et al. Eds.), Greene Publishing Associates & Wiley Interscience, New York (1991); Moss, B., et al., U.S. Patent Number 5,135,855, issued 4 August 1992}, 15 expression in bacteria {Ausubel, F.M., et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley and Sons, Inc., Media PA; Clontech}, expression in yeast {Rosenberg, S. and Tekamp-Olson, P., U.S. Patent No. RE35,749, issued, March 17, 1998; Shuster, J.R., U.S. Patent No. 5,629,203, 20 issued May 13, 1997; Gellissen, G., et al., Antonie Van Leeuwenhoek, 62(1-2):79-93 (1992); Romanos, M.A., et al., Yeast 8(6):423-488 (1992); Goeddel, D.V., Methods in Enzymology 185 (1990); Guthrie, C., and G.R. Fink, Methods in Enzymology 194 (1991)}, expression in 25 mammalian cells {Clontech; Gibco-BRL, Ground Island, NY; e.g., Chinese hamster ovary (CHO) cell lines (Haynes, J., et al., Nuc. Acid. Res. 11:687-706 (1983); 1983, Lau, Y.F., et al., Mol. Cell. Biol. 4:1469-1475 (1984); Kaufman, R. J., "Selection and coamplification of 30 heterologous genes in mammalian cells," in Methods in Enzymology, vol. 185, pp537-566. Academic Press, Inc., San Diego CA (1991)}, and expression in plant cells {plant cloning vectors, Clontech Laboratories, Inc., Palo Alto, CA, and Pharmacia LKB Biotechnology, Inc.,

Piscataway, NJ; Hood, E., et al., *J. Bacteriol.* 168:1291-1301 (1986); Nagel, R., et al., *FEMS Microbiol. Lett.* 67:325 (1990); An, et al., "Binary Vectors", and others in Plant Molecular Biology Manual A3:1-19 (1988);
5 Miki, B.L.A., et al., pp.249-265, and others in Plant DNA Infectious Agents (Hohn, T., et al., eds.) Springer-Verlag, Wien, Austria, (1987); *Plant Molecular Biology: Essential Techniques*, P.G. Jones and J.M. Sutton, New York, J. Wiley, 1997; Miglani, Gurbachan *Dictionary of Plant Genetics and Molecular Biology*, New York, Food Products Press, 1998; Henry, R. J., *Practical Applications of Plant Molecular Biology*, New York, Chapman & Hall, 1997}.

Also included in the invention is an expression
15 vector, such as the CMV promoter-containing vectors described in Example 1, containing coding sequences and expression control elements which allow expression of the coding regions in a suitable host. The control elements generally include a promoter, translation initiation
20 codon, and translation and transcription termination sequences, and an insertion site for introducing the insert into the vector. Translational control elements have been reviewed by M. Kozak (e.g., Kozak, M., *Mamm. Genome* 7(8):563-574, 1996; Kozak, M., *Biochimie* 76(9):815-821, 1994; Kozak, M., *J Cell Biol* 108(2):229-241, 1989; Kozak, M., and Shatkin, A.J.,
25 *Methods Enzymol* 60:360-375, 1979).

Expression in yeast systems has the advantage of commercial production. Recombinant protein production by
30 vaccinia and CHO cell line have the advantage of being mammalian expression systems. Further, vaccinia virus expression has several advantages including the following: (i) its wide host range; (ii) faithful post-

transcriptional modification, processing, folding, transport, secretion, and assembly of recombinant proteins; (iii) high level expression of relatively soluble recombinant proteins; and (iv) a large capacity 5 to accommodate foreign DNA.

The recombinantly expressed polypeptides from synthetic Gag-encoding expression cassettes are typically isolated from lysed cells or culture media. Purification can be carried out by methods known in the art including 10 salt fractionation, ion exchange chromatography, gel filtration, size-exclusion chromatography, size-fractionation, and affinity chromatography. Immunoaffinity chromatography can be employed using antibodies generated based on, for example, Gag antigens.

15 Advantages of expressing the Gag-containing proteins of the present invention using mammalian cells include, but are not limited to, the following: well-established protocols for scale-up production; the ability to produce VLPs; cell lines are suitable to meet good manufacturing 20 process (GMP) standards; culture conditions for mammalian cells are known in the art.

2.1.4 MODIFICATION OF HIV-1 ENV NUCLEIC ACID CODING SEQUENCES

25 One aspect of the present invention is the generation of HIV-1 Env protein coding sequences, and related sequences, having improved expression relative to the corresponding wild-type sequence. Exemplary embodiments of the present invention are illustrated 30 herein modifying the Env protein wild-type sequences obtained from the HIV-1 subtype B strains HIV-1US4 and HIV-1SF162 (Myers et al., Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers et al., *Human Retroviruses and Aids*, 1997, Los Alamos,

New Mexico: Los Alamos National Laboratory). Env sequence obtained from other HIV variants may be manipulated in similar fashion following the teachings of the present specification. Such other variants include those 5 described above in Section 2.1.1 and on the World Wide Web (Internet), for example at <http://hiv-web.lanl.gov/cgi-bin/hivDB3/public/wdb/ssampublic> and <http://hiv-web.lanl.gov>.

First, the HIV-1 codon usage pattern was modified so 10 that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes (Example 1). The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content 15 in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The Env coding sequences were modified to be comparable to codon usage found in highly 20 expressed human genes. Experiments performed in support of the present invention showed that the synthetic Env sequences were capable of higher level of protein production (see the Examples) relative to the native Env sequences. One reason for this increased production may 25 be increased stability of the mRNA corresponding to the synthetic Env coding sequences versus the mRNA corresponding to the native Env coding sequences.

Modification of the Env polypeptide coding sequences resulted in improved expression relative to the wild-type 30 coding sequences in a number of mammalian cell lines. Similar Env polypeptide coding sequences can be obtained from a variety of isolates (families, sub-types, etc.). Env polypeptide encoding sequences derived from these variants can be optimized and tested for improved

expression in mammals by following the teachings of the present specification (see the Examples, in particular Example 2).

5 **2.1.5 FURTHER MODIFICATION OF HIV-1 ENV NUCLEIC ACID
 CODING SEQUENCES**

- In addition to proteins containing HIV-related sequences, the Env encoding sequences of the present invention can be fused to other polypeptides (creating 10 chimeric polypeptides). Also, variations on the orientation of the Env and other coding sequences, relative to each other, are contemplated. Further, the HIV protein encoding cassettes of the present invention can be co-expressed using one vector or multiple vectors.
- 15 In addition, the polyproteins can be operably linked to the same or different promoters.

Env polypeptide coding sequences can be obtained from any HIV isolates (different families, subtypes, and strains) including but not limited to the isolates HIV_{IIIb}, 20 HIV_{SF2}, HIV_{us4}, HIV_{CM235}, HIV_{SF162}, HIV_{LAV}, HIV_{LAI}, HIV_{MN}) (see, e.g., Myers et al., Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers et al., *Human Retroviruses and Aids*, 1997, Los Alamos, New Mexico: Los Alamos National Laboratory). Synthetic 25 expression cassettes can be generated using such coding sequences as starting material by following the teachings of the present specification (e.g., see Example 1). Further, the synthetic expression cassettes (and purified polynucleotides) of the present invention include related 30 Env polypeptide coding sequences having greater than 90%, preferably greater than 92%, more preferably greater than 95%, and most preferably greater than 98% sequence identity to the synthetic expression cassette sequences disclosed herein (for example, SEQ ID NOs:71-72; and/or

the sequences presented in Tables 1A and 1B) when the sequences of the present invention are used as the query sequence.

5 **2.1.6 EXPRESSION OF SYNTHETIC SEQUENCES ENCODING HIV-1
ENV AND RELATED POLYPEPTIDES**

Several synthetic Env-encoding sequences (expression cassettes) of the present invention were cloned into a number of different expression vectors (Example 1) to evaluate levels of expression and production of Env polypeptide. A modified synthetic coding sequence is presented as synthetic Env expression cassettes (Example 1, e.g., Tables 1A and 1B). The synthetic DNA fragments for Env were cloned into eucaryotic expression vectors described in Example 1 and in Section 2.1.3 above, including, a transient expression vector and CMV-promoter-based mammalian vectors. Corresponding wild-type sequences were cloned into the same vectors.

These vectors were then transfected into a several different cell types, including a variety of mammalian cell lines, (293, RD, COS-7, and CHO, cell lines available, for example, from the A.T.C.C.). The cell lines were cultured under appropriate conditions and the levels of gp120, gp140 and gp160 Env expression in supernatants were evaluated (Example 2). Env polypeptides include, but are not limited to, for example, native gp160, oligomeric gp140, monomeric gp120 as well as modified sequences of these polypeptides. The results of these assays demonstrated that expression of synthetic Env encoding sequences were significantly higher than corresponding wild-type sequences (Example 2; Tables 3 and 4).

Further, Western Blot analysis showed that cells containing the synthetic Env expression cassette produced

the expected protein (gp120, gp140 or gp160) at higher per-cell concentrations than cells containing the native expression cassette. The Env proteins were seen in both cell lysates and supernatants. The levels of production
5 were significantly higher in cell supernatants for cells transfected with the synthetic Env expression cassettes of the present invention as compared to wild type.

Fractionation of the supernatants from mammalian cells transfected with the synthetic Env expression
10 cassettes showed that it provides superior production of Env proteins, relative to the wild-type Env sequences (Examples 2 and 3).

Efficient expression of these Env-containing polypeptides in mammalian cell lines provides the
15 following benefits: the Env polypeptides are free of baculovirus or other viral contaminants; production by established methods approved by the FDA; increased purity; greater yields (relative to native coding sequences); and a novel method of producing the Env-
20 containing polypeptides in CHO cells which is less feasible in the absence of the increased expression obtained using the constructs of the present invention.

Exemplary cell lines (e.g., mammalian, yeast, insect, etc.) include those described above in Section
25 2.1.3 for Gag-containing constructs. Further, appropriate vectors and control elements (e.g., promoters, enhancers, polyadenylation sequences, etc.) for any given cell type can be selected, as described above in Section 2.1.3, by one having ordinary skill in the art in view of the
30 teachings of the present specification and information known in the art about expression vectors. In addition, the recombinantly expressed polypeptides from synthetic Env-encoding expression cassettes are typically isolated and purified from lysed cells or culture media, as

described above for Gag-encoding expression cassettes. An exemplary purification is described in Example 4 and shown in Figure 60.

5 **2.1.7 MODIFICATION OF HIV-1 TAT NUCLEIC ACID CODING
SEQUENCES**

Another aspect of the present invention is the generation of HIV-1 tat protein coding sequences, and related sequences, having improved expression relative to 10 the corresponding wild-type sequence. Exemplary embodiments of the present invention are illustrated herein modifying the tat wild-type nucleotide sequence (SEQ ID NO:85, Figure 76) obtained from SF162 as described above. Exemplary synthetic tat constructs are 15 shown in SEQ ID NO:87, which depicts a tat construct encoding a full-length tat polypeptide from strain SF162; SEQ ID NO:88, which depicts a tat construct encoding a tat polypeptide having the cysteine residue at position 22 changed; and SEQ ID NO:89, which depicts a tat construct 20 encoding the amino terminal portion of a tat polypeptide from strain SF162. The amino portion of the tat protein appears to contain many of the epitopes that induce an immune response. In addition, further modifications include replacement or deletion of the cysteine residue at 25 position 22, for example with a valine residue, an alanine residue or a glycine residue (SEQ ID Nos: 88 and 89, Figures 79 and 81), see, e.g., Caputo et al. (1996) *Gene Ther.* 3:235. In Figure 81, which depicts a tat construct encoding the amino terminal portion of a tat 30 polypeptide, the nucleotides (nucleotides 64-66) encoding the cysteine residues are underlined. The design and construction of suitable construct can be readily done using

the teachings of the present specification. As with Gag, pol, prot and Env, tat polypeptide coding sequences can be obtained from a variety of isolates (families, sub-types, etc.).

5 Modification of the tat polypeptide coding sequences result in improved expression relative to the wild-type coding sequences in a number of cell lines (e.g., mammalian, yeast, bacterial and insect cells). Tat polypeptide encoding sequences derived from these
10 variants can be optimized and tested for improved expression in mammals by following the teachings of the present specification (see the Examples, in particular Example 2).

Various forms of the different embodiments of the
15 invention, described herein, may be combined. For example, polynucleotides may be derived from the polynucleotide sequences of the present invention, including, but not limited to, coding sequences for Gag polypeptides, Env polypeptides, polymerase polypeptides,
20 protease polypeptides, tat polypeptides, and reverse transcriptase polypeptides. Further, the polynucleotide coding sequences of the present invention may be combined into multi-cistronic expression cassettes where typically each coding sequence for each polypeptide is preceded by
25 IRES sequences.

**2.2 PRODUCTION OF VIRUS-LIKE PARTICLES AND USE OF THE
CONSTRUCTS OF THE PRESENT INVENTION TO CREATE PACKAGING
CELL LINES**

30 The group-specific antigens (Gag) of human immunodeficiency virus type-1 (HIV-1) self-assemble into noninfectious virus-like particles (VLP) that are released from various eucaryotic cells by budding (reviewed by Freed, E.O., *Virology* 251:1-15, 1998). The

synthetic expression cassettes of the present invention provide efficient means for the production of HIV-Gag virus-like particles (VLPs) using a variety of different cell types, including, but not limited to, mammalian cells.

5 Viral particles can be used as a matrix for the proper presentation of an antigen entrapped or associated therewith to the immune system of the host. For example, U.S. Patent No. 4,722,840 describes hybrid particles comprised of a particle-forming fragment of a structural protein from a virus, such as a particle-forming fragment of hepatitis B virus (HBV) surface antigen (HBsAg), fused to a heterologous polypeptide. Tindle et al., *Virology* 10 (1994) 200:547-557, describes the production and use of chimeric HBV core antigen particles containing epitopes of human papillomavirus (HPV) type 16 E7 transforming protein.

15 Adams et al., *Nature* (1987) 329:68-70, describes the recombinant production of hybrid HIVgp120:Ty VLPs in yeast and Brown et al., *Virology* (1994) 198:477-488, the production of chimeric proteins consisting of the VP2 protein of human parvovirus B19 and epitopes from human herpes simplex virus type 1, as well as mouse hepatitis virus A59. Wagner et al., (*Virology* (1994) 200:162-175, 20 Brand et al., *J. Virol. Meth.* (1995) 51:153-168; *Virology* (1996) 220:128-140) and Wolf, et al., (EP 0 449 116 A1, published 2 October 1991; WO 96/30523, published 3 October 1996) describe the assembly of chimeric HIV-1 p55Gag particles. U.S. Patent No. 5,503,833 describes 25 the use of rotavirus VP6 spheres for encapsulating and 30 delivering therapeutic agents.

**2.2.1 VLP PRODUCTION USING THE SYNTHETIC EXPRESSION
CASSETTES OF THE PRESENT INVENTION**

Experiments performed in support of the present invention have demonstrated that the synthetic expression 5 cassettes of the present invention provide superior production of both protein and VLPs, relative to native coding sequences (Examples 7 and 15). Further, electron microscopic evaluation of VLP production (Examples 6 and 10, Figures 3A-B and 65A-F) showed that free and budding immature virus particles of the expected size were produced by cells containing the synthetic expression cassettes.

Using the synthetic expression cassettes of the present invention, rather than native coding sequences, 15 for the production of virus-like particles provide several advantages. First, VLPs can be produced in enhanced quantity making isolation and purification of the VLPs easier. Second, VLPs can be produced in a variety of cell types using the synthetic expression 20 cassettes, in particular, mammalian cell lines can be used for VLP production, for example, CHO cells. Production using CHO cells provides (i) VLP formation; (ii) correct myristylation and budding; (iii) absence of non-mammalian cell contaminants (e.g., insect viruses 25 and/or cells); and (iv) ease of purification. The synthetic expression cassettes of the present invention are also useful for enhanced expression in cell-types other than mammalian cell lines. For example, infection of insect cells with baculovirus vectors encoding the 30 synthetic expression cassettes resulted in higher levels of total protein yield and higher levels of VLP production (relative to wild-type coding sequences). Further, the final product from insect cells infected with the baculovirus-Gag synthetic expression cassettes

consistently contained lower amounts of contaminating insect proteins than the final product when wild-type coding sequences were used (Examples).

VLPs can spontaneously form when the particle-forming polypeptide of interest is recombinantly expressed in an appropriate host cell. Thus, the VLPs produced using the synthetic expression cassettes of the present invention are conveniently prepared using recombinant techniques. As discussed below, the Gag polypeptide encoding synthetic expression cassettes of the present invention can include other polypeptide coding sequences of interest (for example, Env, tat, rev, HIV protease, HIV polymerase, HCV core; see, Example 1). Expression of such synthetic expression cassettes yields VLPs comprising the product of the synthetic expression cassette, as well as, the polypeptide of interest.

Once coding sequences for the desired particle-forming polypeptides have been isolated or synthesized, they can be cloned into any suitable vector or replicon for expression. Numerous cloning vectors are known to those of skill in the art, and the selection of an appropriate cloning vector is a matter of choice. See, generally, Ausubel et al, *supra* or Sambrook et al, *supra*. The vector is then used to transform an appropriate host cell. Suitable recombinant expression systems include, but are not limited to, bacterial, mammalian, baculovirus/insect, vaccinia, Semliki Forest virus (SFV), Alphaviruses (such as, Sindbis, Venezuelan Equine Encephalitis (VEE)), mammalian, yeast and *Xenopus* expression systems, well known in the art. Particularly preferred expression systems are mammalian cell lines, vaccinia, Sindbis, insect and yeast systems.

For example, a number of mammalian cell lines are known in the art and include immortalized cell lines

available from the American Type Culture Collection (A.T.C.C.), such as, but not limited to, Chinese hamster ovary (CHO) cells, 293 cells, HeLa cells, baby hamster kidney (BHK) cells, mouse myeloma (SB20), monkey kidney cells (COS), as well as others. Similarly, bacterial hosts such as *E. coli*, *Bacillus subtilis*, and *Streptococcus spp.*, will find use with the present expression constructs. Yeast hosts useful in the present invention include *inter alia*, *Saccharomyces cerevisiae*, 5 *Candida albicans*, *Candida maltosa*, *Hansenula polymorpha*, *Kluyveromyces fragilis*, *Kluyveromyces lactis*, *Pichia guillermondii*, *Pichia pastoris*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Insect cells for use with baculovirus expression vectors include, *inter alia*, *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni*. 10 See, e.g., Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987). Fungal hosts include, for example, *Aspergillus*.

15 Viral vectors can be used for the production of particles in eucaryotic cells, such as those derived from the pox family of viruses, including vaccinia virus and avian poxvirus. Additionally, a vaccinia based infection/transfection system, as described in Tomei et 20 al., *J. Virol.* (1993) 67:4017-4026 and Selby et al., *J. Gen. Virol.* (1993) 74:1103-1113, will also find use with the present invention. In this system, cells are first infected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This 25 polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the DNA of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus 30

recombinant transcribes the transfected DNA into RNA which is then translated into protein by the host translational machinery. Alternately, T7 can be added as a purified protein or enzyme as in the "Progenitor" system (Studier and Moffatt, *J. Mol. Biol.* (1986) 189:113-130). The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation product(s).

Depending on the expression system and host selected, the VLPs are produced by growing host cells transformed by an expression vector under conditions whereby the particle-forming polypeptide is expressed and VLPs can be formed. The selection of the appropriate growth conditions is within the skill of the art. If the VLPs are formed intracellularly, the cells are then disrupted, using chemical, physical or mechanical means, which lyse the cells yet keep the VLPs substantially intact. Such methods are known to those of skill in the art and are described in, e.g., *Protein Purification Applications: A Practical Approach*, (E.L.V. Harris and S. Angal, Eds., 1990).

The particles are then isolated (or substantially purified) using methods that preserve the integrity thereof, such as, by density gradient centrifugation, e.g., sucrose gradients, PEG-precipitation, pelleting, and the like (see, e.g., Kirnbauer et al. *J. Virol.* (1993) 67:6929-6936), as well as standard purification techniques including, e.g., ion exchange and gel filtration chromatography.

VLPs produced by cells containing the synthetic expression cassettes of the present invention can be used to elicit an immune response when administered to a subject. One advantage of the present invention is that VLPs can be produced by mammalian cells carrying the

synthetic expression cassettes at levels previously not possible. As discussed above, the VLPs can comprise a variety of antigens in addition to the Gag polypeptides (e.g., Env, tat, Gag-protease, Gag-polymerase, Gag-HCV-core). Purified VLPs, produced using the synthetic expression cassettes of the present invention, can be administered to a vertebrate subject, usually in the form of vaccine compositions. Combination vaccines may also be used, where such vaccines contain, for example, other subunit proteins derived from HIV or other organisms (e.g., env) or gene delivery vaccines encoding such antigens. Administration can take place using the VLPs formulated alone or formulated with other antigens. Further, the VLPs can be administered prior to, concurrent with, or subsequent to, delivery of the synthetic expression cassettes for DNA immunization (see below) and/or delivery of other vaccines. Also, the site of VLP administration may be the same or different as other vaccine compositions that are being administered.

Gene delivery can be accomplished by a number of methods including, but are not limited to, immunization with DNA, alphavirus vectors, pox virus vectors, and vaccinia virus vectors.

VLP immune-stimulating (or vaccine) compositions can include various excipients, adjuvants, carriers, auxiliary substances, modulating agents, and the like. The immune stimulating compositions will include an amount of the VLP/antigen sufficient to mount an immunological response. An appropriate effective amount can be determined by one of skill in the art. Such an amount will fall in a relatively broad range that can be determined through routine trials and will generally be an amount on the order of about 0.1 μ g to about 1000 μ g,

more preferably about 1 μg to about 300 μg , of VLP/antigen.

A carrier is optionally present which is a molecule that does not itself induce the production of antibodies 5 harmful to the individual receiving the composition.

Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as 10 oil droplets or liposomes), and inactive virus particles. Examples of particulate carriers include those derived from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-glycolides), known as PLG. See, e.g., 15 Jeffery et al., *Pharm. Res.* (1993) 10:362-368; McGee JP, et al., *J Microencapsul.* 14(2):197-210, 1997; O'Hagan DT, et al., *Vaccine* 11(2):149-54, 1993. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as 20 immunostimulating agents ("adjuvants"). Furthermore, the antigen may be conjugated to a bacterial toxoid, such as toxoid from diphtheria, tetanus, cholera, etc., as well as toxins derived from *E. coli*.

Such adjuvants include, but are not limited to: (1) 25 aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc.; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), 30 such as for example (a) MF59 (International Publication No. WO 90/14837), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated

into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below)

5 either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group

10 consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particle generated therefrom such as

15 ISCOMs (immunostimulating complexes); (4) Complete Freunds Adjuvant (CFA) and Incomplete Freunds Adjuvant (IFA); (5) cytokines, such as interleukins (IL-1, IL-2, etc.), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), beta chemokines (MIP, 1-

20 alpha, 1-beta Rantes, etc.); (6) detoxified mutants of a bacterial ADP-ribosylating toxin such as a cholera toxin (CT), a pertussis toxin (PT), or an *E. coli* heat-labile toxin (LT), particularly LT-K63 (where lysine is substituted for the wild-type amino acid at position 63)

25 LT-R72 (where arginine is substituted for the wild-type amino acid at position 72), CT-S109 (where serine is substituted for the wild-type amino acid at position 109), and PT-K9/G129 (where lysine is substituted for the wild-type amino acid at position 9 and glycine

30 substituted at position 129) (see, e.g., International Publication Nos. W093/13202 and W092/19265); and (7)

other substances that act as immunostimulating agents to enhance the effectiveness of the composition.

Muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-
5 acteyl-normuramyl-L-alanyl-D-isogluatme (nor-MDP), N-
acetylmuramyl-L-alanyl-D-isogluatminyl-L-alanine-2-(1'-
2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-
ethylamine (MTP-PE), etc.

Dosage treatment with the VLP composition may be a
10 single dose schedule or a multiple dose schedule. A
multiple dose schedule is one in which a primary course
of vaccination may be with 1-10 separate doses, followed
by other doses given at subsequent time intervals, chosen
to maintain and/or reinforce the immune response, for
15 example at 1-4 months for a second dose, and if needed, a
subsequent dose(s) after several months. The dosage
regimen will also, at least in part, be determined by the
potency of the modality, the vaccine delivery employed,
the need of the subject and be dependent on the judgment
20 of the practitioner.

If prevention of disease is desired (e.g., reduction
of symptoms, recurrences or of disease progression), the
antigen carrying VLPs are generally administered prior to
primary infection with the pathogen of interest. If
25 treatment is desired, e.g., the reduction of symptoms or
recurrences, the VLP compositions are generally
administered subsequent to primary infection.

2.2.2 **USING THE SYNTHETIC EXPRESSION CASSETTES OF THE
30 PRESENT INVENTION TO CREATE PACKAGING CELL LINES**

A number of viral based systems have been developed
for use as gene transfer vectors for mammalian host
cells. For example, retroviruses (in particular,

lentiviral vectors) provide a convenient platform for gene delivery systems. A coding sequence of interest (for example, a sequence useful for gene therapy applications) can be inserted into a gene delivery vector and packaged in retroviral particles using techniques known in the art. Recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of retroviral systems have been described, including, for example, the following: (U.S. Patent No. 5,219,740; Miller et al. (1989) *Biotechniques* 7:980; Miller, A.D. (1990) *Human Gene Therapy* 1:5; Scarpa et al. (1991) *Virology* 180:849; Burns et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:8033; Boris-Lawrie et al. (1993) *Cur. Opin. Genet. Develop.* 3:102; GB 2200651; EP 0415731; EP 0345242; WO 89/02468; WO 89/05349; WO 89/09271; WO 90/02806; WO 90/07936; WO 90/07936; WO 94/03622; WO 93/25698; WO 93/25234; WO 93/11230; WO 93/10218; WO 91/02805; in U.S. 5,219,740; U.S. 4,405,712; U.S. 4,861,719; U.S. 4,980,289 and U.S. 4,777,127; in U.S. Serial No. 07/800,921; and in Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53:83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci USA* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Sequences useful for gene therapy applications include, but are not limited to, the following. Factor VIII cDNA, including derivatives and deletions thereof (International Publication Nos. WO 96/21035, WO 97/03193, WO 97/03194, WO 97/03195, and WO 97/03191). Factor IX cDNA (Kurachi et al. (1982) *Proc. Natl. Acad. Sci. USA* 79:6461-6464). Factor V cDNA can be obtained from pMT2-V (Jenny (1987) *Proc. Natl. Acad. Sci. USA* 84:4846, A.T.C.C. Deposit No. 40515). A full-length factor V

cDNA, or a B domain deletion or B domain substitution thereof, can be used. B domain deletions of factor V, include those reported by Marquette (1995) *Blood* 86:3026 and Kane (1990) *Biochemistry* 29:6762. Antithrombin III cDNA (Prochownik (1983) *J. Biol. Chem.* 258:8389, A.T.C.C. Deposit No. 57224/57225). Protein C encoding cDNA (Foster (1984) *Proc. Natl. Acad. Sci. USA* 81:4766; Beckmann (1985) *Nucleic Acids Res.* 13:5233). Prothrombin cDNA can be obtained by restriction enzyme digestion of a published vector (Degen (1983) *Biochemistry* 22:2087). The endothelial cell surface protein, thrombomodulin, is a necessary cofactor for the normal activation of protein C by thrombin. A soluble recombinant form has been described (Parkinson (1990) *J. Biol. Chem.* 265:12602; Jackman (1987) *Proc. Natl. Acad. Sci. USA* 84:6425; Shirai (1988) *J. Biochem.* 103:281; Wen (1987) *Biochemistry* 26:4350; Suzuki (1987) *EMBO J.* 6:1891, A.T.C.C. Deposit No. 61348, 61349).

Many genetic diseases caused by inheritance of defective genes result in the failure to produce normal gene products, for example, thalassemia, phenylketonuria, Lesch-Nyhan syndrome, severe combined immunodeficiency (SCID), hemophilia A and B, cystic fibrosis, Duchenne's Muscular Dystrophy, inherited emphysema and familial hypercholesterolemia (Mulligan et al. (1993) *Science* 260:926; Anderson et al. (1992) *Science* 256:808; Friedman et al. (1989) *Science* 244:1275). Although genetic diseases may result in the absence of a gene product, endocrine disorders, such as diabetes and hypopituitarism, are caused by the inability of the gene to produce adequate levels of the appropriate hormone insulin and human growth hormone respectively.

In one aspect, gene therapy employing the constructs and methods of the present invention involves the

introduction of normal recombinant genes into T cells so that new or missing proteins are produced by the T cells after introduction or reintroduction thereof into a patient. A number of genetic diseases have been selected
5 for treatment with gene therapy, including adenine deaminase deficiency, cystic fibrosis, α_1 -antitrypsin deficiency, Gaucher's syndrome, as well as non-genetic diseases.

In particular, Gaucher's syndrome is a genetic
10 disorder characterized by a deficiency of the enzyme glucocerebrosidase. This enzyme deficiency leads to the accumulation of glucocerebroside in the lysosomes of all cells in the body. For a review see *Science* 256:794 (1992) and Scriver et al., *The Metabolic Basis of
15 Inherited Disease*, 6th ed., vol. 2, page 1677). Thus, gene transfer vectors that express glucocerebrosidase can be constructed for use in the treatment of this disorder. Likewise, gene transfer vectors encoding lactase can be used in the treatment of hereditary lactose intolerance,
20 those expressing AD can be used for treatment of ADA deficiency, and gene transfer vectors encoding α_1 -antitrypsin can be used to treat α_1 -antitrypsin deficiency. See Ledley, F.D. (1987) *J. Pediatrics* 110:157-174, Verma, I. (Nov. 1987) *Scientific American* pp. 68-84, and International Publication No. WO 95/27512
25 entitled "Gene Therapy Treatment for a Variety of Diseases and Disorders," for a description of gene therapy treatment of genetic diseases.

In still further embodiments of the invention,
30 nucleotide sequences which can be incorporated into a gene transfer vector include, but are not limited to, proteins associated with enzyme-deficiency disorders, such as the cystic fibrosis transmembrane regulator (see, for example, U.S. Patent No. 5,240,846 and Larrick et al.

(1991) *Gene Therapy Applications of Molecular Biology*, Elsevier, New York and adenosine deaminase (ADA) (see U.S. Patent No. 5,399,346); growth factors, or an agonist or antagonist of a growth factor (Bandara et al. (1992) 5 *DNA and Cell Biology*, 11:227); one or more tumor suppressor genes such as p53, Rb, or C-CAMI (Kleinerman et al. (1995) *Cancer Research* 55:2831); a molecule that modulates the immune system of an organism, such as a HLA molecule (Nabel et al. (1993) *Proc. Natl. Acad. Sci. USA* 10 90:11307); a ribozyme (Larsson et al. (1996) *Virology* 219:161); a peptide nucleic acid (Hirshman et al. (1996) *J. Invest. Med.* 44:347); an antisense molecule (Bordier et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:9383) which can be used to down-regulate the expression or synthesis 15 of aberrant or foreign proteins, such as HIV proteins or a wide variety of oncogenes such as p53 (Hesketh, *The Oncogene Facts Book*, Academic Press, New York, (1995); a biopharmaceutical agent or antisense molecule used to treat HIV-infection, such as an inhibitor of p24 20 (Nakashima et al. (1994) *Nucleic Acids Res.* 22:5004); or reverse-transcriptase (see, Bordier, *supra*).

Other proteins of therapeutic interest can be expressed *in vivo* by gene transfer vectors using the methods of the invention. For instance sustained *in vivo* 25 expression of tissue factor inhibitory protein (TFPI) is useful for treatment of conditions including sepsis and DIC and in preventing reperfusion injury. (See International Publications Nos. WO 93/24143, WO 93/25230 and WO 96/06637). Nucleic acid sequences encoding 30 various forms of TFPI can be obtained, for example, as described in US Patent Nos. 4,966,852; 5,106,833; and 5,466,783, and incorporated into the gene transfer vectors described herein.

Erythropoietin (EPO) and leptin can also be expressed *in vivo* from genetically modified T cells according to the methods of the invention. For instance EPO is useful in gene therapy treatment of a variety of disorders including anemia (see International Publication No. WO 95/13376 entitled "Gene Therapy for Treatment of Anemia"). Sustained delivery of leptin by the methods of the invention is useful in treatment of obesity. See International Publication No. WO 96/05309 for a description of the leptin gene and the use thereof in the treatment of obesity.

A variety of other disorders can also be treated by the methods of the invention. For example, sustained *in vivo* systemic production of apolipoprotein E or apolipoprotein A from genetically modified T cells can be used for treatment of hyperlipidemia (see Breslow et al. (1994) *Biotechnology* 12:365). Sustained production of angiotensin receptor inhibitor (Goodfriend et al. (1996) *N. Engl. J. Med.* 334:1469) can be provided by the methods described herein. As yet an additional example, the long term *in vivo* systemic production of angiostatin is useful in the treatment of a variety of tumors. (See O'Reilly et al. (1996) *Nature Med.* 2:689).

In other embodiments, gene transfer vectors can be constructed to encode a cytokine or other immunomodulatory molecule. For example, nucleic acid sequences encoding native IL-2 and gamma-interferon can be obtained as described in US Patent Nos. 4,738,927 and 5,326,859, respectively, while useful muteins of these proteins can be obtained as described in U.S. Patent No. 4,853,332. Nucleic acid sequences encoding the short and long forms of mCSF can be obtained as described in US Patent Nos. 4,847,201 and 4,879,227, respectively. In particular aspects of the invention, retroviral vectors

expressing cytokine or immunomodulatory genes can be produced as described herein (for example, employing the packaging cell lines of the present invention) and in International Application No. PCT US 94/02951, entitled "Compositions and Methods for Cancer Immunotherapy."

Examples of suitable immunomodulatory molecules for use herein include the following: IL-1 and IL-2 (Karupiah et al. (1990) *J. Immunology* 144:290-298, Weber et al. (1987) *J. Exp. Med.* 166:1716-1733, Gansbacher et al. (1990) *J. Exp. Med.* 172:1217-1224, and U.S. Patent No. 4,738,927); IL-3 and IL-4 (Tepper et al. (1989) *Cell* 57:503-512, Golumbek et al. (1991) *Science* 254:713-716, and U.S. Patent No. 5,017,691); IL-5 and IL-6 (Brakenhof et al. (1987) *J. Immunol.* 139:4116-4121, and 15 International Publication No. WO 90/06370); IL-7 (U.S. Patent No. 4,965,195); IL-8, IL-9, IL-10, IL-11, IL-12, and IL-13 (*Cytokine Bulletin*, Summer 1994); IL-14 and IL-15; alpha interferon (Finter et al. (1991) *Drugs* 42:749-765, U.S. Patent Nos. 4,892,743 and 4,966,843, 20 International Publication No. WO 85/02862, Nagata et al. (1980) *Nature* 284:316-320, Familletti et al. (1981) *Methods in Enz.* 78:387-394, Twu et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:2046-2050, and Faktor et al. (1990) *Oncogene* 5:867-872); beta-interferon (Seif et al. (1991) *J. Virol.* 65:664-671); gamma-interferons (Radford et al. 25 (1991) *The American Society of Hepatology* 20082015, Watanabe et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:9456-9460, Gansbacher et al. (1990) *Cancer Research* 50:7820-7825, Maio et al. (1989) *Can. Immunol.* 30 *Immunother.* 30:34-42, and U.S. Patent Nos. 4,762,791 and 4,727,138); G-CSF (U.S. Patent Nos. 4,999,291 and 4,810,643); GM-CSF (International Publication No. WO 85/04188); tumor necrosis factors (TNFs) (Jayaraman et al. (1990) *J. Immunology* 144:942-951); CD3 (Krissanen et

al. (1987) *Immunogenetics* 26:258-266); ICAM-1 (Altman et al. (1989) *Nature* 338:512-514, Simmons et al. (1988) *Nature* 331:624-627); ICAM-2, LFA-1, LFA-3 (Wallner et al. (1987) *J. Exp. Med.* 166:923-932); MHC class I molecules, 5 MHC class II molecules, B7.1-.3, β_2 -microglobulin (Parnes et al. (1981) *Proc. Natl. Acad. Sci. USA* 78:2253-2257); chaperones such as calnexin; and MHC-linked transporter proteins or analogs thereof (Powis et al. (1991) *Nature* 354:528-531). Immunomodulatory factors may also be 10 agonists, antagonists, or ligands for these molecules. For example, soluble forms of receptors can often behave as antagonists for these types of factors, as can mutated forms of the factors themselves.

Nucleic acid molecules that encode the above- 15 described substances, as well as other nucleic acid molecules that are advantageous for use within the present invention, may be readily obtained from a variety of sources, including, for example, depositories such as the American Type Culture Collection, or from commercial 20 sources such as British Bio-Technology Limited (Cowley, Oxford England). Representative examples include BBG 12 (containing the GM-CSF gene coding for the mature protein of 127 amino acids), BBG 6 (which contains sequences encoding gamma interferon), A.T.C.C. Deposit No. 39656 25 (which contains sequences encoding TNF), A.T.C.C. Deposit No. 20663 (which contains sequences encoding alpha- interferon), A.T.C.C. Deposit Nos. 31902, 31902 and 39517 (which contain sequences encoding beta-interferon), A.T.C.C. Deposit No. 67024 (which contains a sequence 30 which encodes Interleukin-1b), A.T.C.C. Deposit Nos. 39405, 39452, 39516, 39626 and 39673 (which contain sequences encoding Interleukin-2), A.T.C.C. Deposit Nos. 59399, 59398, and 67326 (which contain sequences encoding Interleukin-3), A.T.C.C. Deposit No. 57592 (which

contains sequences encoding Interleukin-4), A.T.C.C. Deposit Nos. 59394 and 59395 (which contain sequences encoding Interleukin-5), and A.T.C.C. Deposit No. 67153 (which contains sequences encoding Interleukin-6).

5 Plasmids containing cytokine genes or immunomodulatory genes (International Publication Nos. WO 94/02951 and WO 96/21015) can be digested with appropriate restriction enzymes, and DNA fragments containing the particular gene of interest can be inserted into a gene
10 transfer vector using standard molecular biology techniques. (See, e.g., Sambrook et al., *supra.*, or Ausubel et al. (eds) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience).

Exemplary hormones, growth factors and other
15 proteins which are useful for long term expression are described, for example, in European Publication No. 0437478B1, entitled "Cyclodextrin-Peptide Complexes." Nucleic acid sequences encoding a variety of hormones can be used, including those encoding human growth hormone,
20 insulin, calcitonin, prolactin, follicle stimulating hormone (FSH), luteinizing hormone (LH), human chorionic gonadotropin (HCG), and thyroid stimulating hormone (TSH). A variety of different forms of IGF-1 and IGF-2 growth factor polypeptides are also well known the art
25 and can be incorporated into gene transfer vectors for long term expression *in vivo*. See, e.g., European Patent No. 0123228B1, published for grant September 19, 1993, entitled "Hybrid DNA Synthesis of Mature Insulin-like Growth Factors." As an additional example, the long term
30 *in vivo* expression of different forms of fibroblast growth factor can also be effected employing the compositions and methods of invention. See, e.g., U.S. Patent Nos. 5,464,774, 5,155,214, and 4,994,559 for a description of different fibroblast growth factors.

Polynucleotide sequences coding for the above-described molecules can be obtained using recombinant methods, such as by screening cDNA and genomic libraries from cells expressing the gene, or by deriving the gene 5 from a vector known to include the same. For example, plasmids which contain sequences that encode altered cellular products may be obtained from a depository such as the A.T.C.C., or from commercial sources. Plasmids containing the nucleotide sequences of interest can be 10 digested with appropriate restriction enzymes, and DNA fragments containing the nucleotide sequences can be inserted into a gene transfer vector using standard molecular biology techniques.

Alternatively, cDNA sequences for use with the 15 present invention may be obtained from cells which express or contain the sequences, using standard techniques, such as phenol extraction and PCR of cDNA or genomic DNA. See, e.g., Sambrook et al., *supra*, for a description of techniques used to obtain and isolate DNA. 20 Briefly, mRNA from a cell which expresses the gene of interest can be reverse transcribed with reverse transcriptase using oligo-dT or random primers. The single stranded cDNA may then be amplified by PCR (see U.S. Patent Nos. 4,683,202, 4,683,195 and 4,800,159, see 25 also *PCR Technology: Principles and Applications for DNA Amplification*, Erlich (ed.), Stockton Press, 1989)) using oligonucleotide primers complementary to sequences on either side of desired sequences.

The nucleotide sequence of interest can also be 30 produced synthetically, rather than cloned, using a DNA synthesizer (e.g., an Applied Biosystems Model 392 DNA Synthesizer, available from ABI, Foster City, California). The nucleotide sequence can be designed with the appropriate codons for the expression product

desired. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, e.g., Edge (1981) *Nature* 292:756; Nambair et al. (1984) 5 *Science* 223:1299; Jay et al. (1984) *J. Biol. Chem.* 259:6311.

The synthetic expression cassettes of the present invention can be employed in the construction of packaging cell lines for use with retroviral vectors.

10 One type of retrovirus, the murine leukemia virus, or "MLV", has been widely utilized for gene therapy applications (see generally Mann et al. (*Cell* 33:153, 1993), Cane and Mulligan (*Proc, Nat'l. Acad. Sci. USA* 81:6349, 1984), and Miller et al., *Human Gene Therapy* 15 1:5-14, 1990).

Lentiviral vectors typically, comprise a 5' lentiviral LTR, a tRNA binding site, a packaging signal, a promoter operably linked to one or more genes of interest, an origin of second strand DNA synthesis and a 20 3' lentiviral LTR, wherein the lentiviral vector contains a nuclear transport element. The nuclear transport element may be located either upstream (5') or downstream (3') of a coding sequence of interest. Within certain embodiments, the nuclear transport element is not RRE. 25 Within one embodiment the packaging signal is an extended packaging signal. Within other embodiments the promoter is a tissue specific promoter, or, alternatively, a promoter such as CMV. Within other embodiments, the lentiviral vector further comprises an internal ribosome 30 entry site.

A wide variety of lentiviruses may be utilized within the context of the present invention, including for example, lentiviruses selected from the group consisting of HIV, HIV-1, HIV-2, FIV and SIV.

In one embodiment of the present invention synthetic Env and/or Gag-polymerase expression cassettes are provided comprising a promoter and a sequence encoding synthetic Gag-polymerase (SEQ ID NO:6) and at least one of vpr, vpu, nef or vif, wherein the promoter is operably linked to Gag-polymerase and vpr, vpu, nef or vif.

Within yet another aspect of the invention, host cells (e.g., packaging cell lines) are provided which contain any of the expression cassettes described herein.

For example, within one aspect packaging cell line are provided comprising an expression cassette that comprises a sequence encoding synthetic Env and/or Gag-polymerase, and a nuclear transport element, wherein the promoter is operably linked to the sequence encoding Env and/or Gag-polymerase. Packaging cell lines may further comprise a promoter and a sequence encoding tat, rev, or an envelope, wherein the promoter is operably linked to the sequence encoding tat, rev, or, the envelope. The packaging cell line may further comprise a sequence encoding any one or more of nef, vif, vpu or vpr.

In one embodiment, the expression cassette (carrying, for example, the synthetic Env, synthetic tat and/or synthetic Gag-polymerase) is stably integrated. The packaging cell line, upon introduction of a lentiviral vector, typically produces viral particles. The promoter regulating expression of the synthetic expression cassette may be inducible. Typically, the packaging cell line, upon introduction of a lentiviral vector, produces viral particles that are essentially free of replication competent virus.

Packaging cell lines are provided comprising an expression cassette which directs the expression of a synthetic Env (or Gag-polymerase) gene, an expression cassette which directs the expression of a Gag (or Env).

gene optimized for expression (e.g., Andre, S., et al., *Journal of Virology* 72(2):1497-1503, 1998; Haas, J., et al., *Current Biology* 6(3):315-324, 1996). A lentiviral vector is introduced into the packaging cell line to
5 produce a vector particle producing cell line.

As noted above, lentiviral vectors can be designed to carry or express a selected gene(s) or sequences of interest. Lentiviral vectors may be readily constructed from a wide variety of lentiviruses (see RNA Tumor
10 Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985). Representative examples of lentiviruses included HIV, HIV-1, HIV-2, FIV and SIV. Such lentiviruses may either be obtained from patient isolates, or, more
15 preferably, from depositories or collections such as the American Type Culture Collection, or isolated from known sources using available techniques.

Portions of the lentiviral gene delivery vectors (or vehicles) may be derived from different viruses. For example, in a given recombinant lentiviral vector, LTRs
20 may be derived from an HIV, a packaging signal from SIV, and an origin of second strand synthesis from HrV-2. Lentiviral vector constructs may comprise a 5' lentiviral LTR, a tRNA binding site, a packaging signal, one or more heterologous sequences, an origin of second strand DNA
25 synthesis and a 3' LTR, wherein said lentiviral vector contains a nuclear transport element that is not RRE.

Briefly, Long Terminal Repeats ("LTRs") are subdivided into three elements, designated U5, R and U3. These elements contain a variety of signals which are
30 responsible for the biological activity of a retrovirus, including for example, promoter and enhancer elements which are located within U3. LTRs may be readily identified in the provirus (integrated DNA form) due to their precise duplication at either end of the genome.

As utilized herein, a 5' LTR should be understood to include a 5' promoter element and sufficient LTR sequence to allow reverse transcription and integration of the DNA form of the vector. The 3' LTR should be understood to 5 include a polyadenylation signal, and sufficient LTR sequence to allow reverse transcription and integration of the DNA form of the vector.

The tRNA binding site and origin of second strand DNA synthesis are also important for a retrovirus to be 10 biologically active, and may be readily identified by one of skill in the art. For example, retroviral tRNA binds to a tRNA binding site by Watson-Crick base pairing, and is carried with the retrovirus genome into a viral particle. The tRNA is then utilized as a primer for DNA 15 synthesis by reverse transcriptase. The tRNA binding site may be readily identified based upon its location just downstream from the 5'LTR. Similarly, the origin of second strand DNA synthesis is, as its name implies, important for the second strand DNA synthesis of a 20 retrovirus. This region, which is also referred to as the poly-purine tract, is located just upstream of the 3'LTR.

In addition to a 5' and 3' LTR, tRNA binding site, and origin of second strand DNA synthesis, recombinant 25 retroviral vector constructs may also comprise a packaging signal, as well as one or more genes or coding sequences of interest. In addition, the lentiviral vectors have a nuclear transport element which, in preferred embodiments is not RRE. Representative examples of suitable nuclear transport elements include 30 the element in Rous sarcoma virus (Ogert, et al., J ViroL 70, 3834-3843, 1996), the element in Rous sarcoma virus (Liu & Mertz, Genes & Dev., 9, 1766-1789, 1995) and the element in the genome of simian retrovirus type I

(Zolotukhin, et al., *J Virol.* 68, 7944-7952, 1994). Other potential elements include the elements in the histone gene (Kedes, *Annu. Rev. Biochem.* 48, 837-870, 1970), the α -interferon gene (Nagata et al., *Nature* 287, 5 401-408, 1980), the β -adrenergic receptor gene (Koilkka, et al., *Nature* 329, 75-79, 1987), and the c-Jun gene (Hattorie, et al., *Proc. Natl. Acad. Sci. USA* 85, 9148-9152, 1988).

Recombinant lentiviral vector constructs typically 10 lack both *Gag-polymerase* and *env* coding sequences. Recombinant lentiviral vector typically contain less than 20, preferably 15, more preferably 10, and most preferably 8 consecutive nucleotides found in *Gag-polymerase* or *env* genes. One advantage of the present 15 invention is that the synthetic *Gag-polymerase* expression cassettes, which can be used to construct packaging cell lines for the recombinant retroviral vector constructs, have little homology to wild-type *Gag-polymerase* sequences and thus considerably reduce or eliminate the 20 possibility of homologous recombination between the synthetic and wild-type sequences.

Lentiviral vectors may also include tissue-specific 25 promoters to drive expression of one or more genes or sequences of interest. For example, lentiviral vector particles of the invention can contain a liver specific promoter to maximize the potential for liver specific expression of the exogenous DNA sequence contained in the vectors. Preferred liver specific promoters include the hepatitis B X-gene promoter and the hepatitis B core 30 protein promoter. These liver specific promoters are preferably employed with their respective enhancers. The enhancer element can be linked at either the 5' or the 3' end of the nucleic acid encoding the sequences of interest. The hepatitis B X gene promoter and its

enhancer can be obtained from the viral genome as a 332 base pair *EcoRV-NcoI* DNA fragment employing the methods described in Twu, et al., *J Virol.* 61:3448-3453, 1987. The hepatitis B core protein promoter can be obtained 5 from the viral genome as a 584 base pair *BamHI-BglII* DNA fragment employing the methods described in Gerlach, et al., *Virol* 189:59-66, 1992. It may be necessary to remove the negative regulatory sequence in the *BamHI-BglII* fragment prior to inserting it. Other liver 10 specific promoters include the AFP (alpha fetal protein) gene promoter and the albumin gene promoter, as disclosed in EP Patent Publication 0 415 731, the -1 antitrypsin gene promoter, as disclosed in Rettenger, et al., *Proc. Natl. Acad. Sci.* 91:1460-1464, 1994, the fibrinogen 15 gene promoter, the APO-A1 (Apolipoprotein A1) gene promoter, and the promoter genes for liver transference enzymes such as, for example, SGOT, SGPT and glutamyle transferase. See also PCT Patent Publications WO 90/07936 and WO 91/02805 for a description of the use of 20 liver specific promoters in lentiviral vector particles.

Lentiviral vector constructs may be generated such that more than one gene of interest is expressed. This may be accomplished through the use of di- or oligo-cistronic cassettes (e.g., where the coding regions are 25 separated by 80 nucleotides or less, see generally Levin et al., *Gene* 108:167-174, 1991), or through the use of Internal Ribosome Entry Sites ("IRES").

Packaging cell lines suitable for use with the above described recombinant retroviral vector constructs may be 30 readily prepared given the disclosure provided herein. Briefly, the parent cell line from which the packaging cell line is derived can be selected from a variety of

mammalian cell lines, including for example, 293, RD, COS-7, CHO, BHK, VERO, HT1080, and myeloma cells.

After selection of a suitable host cell for the generation of a packaging cell line, one or more 5 expression cassettes are introduced into the cell line in order to complement or supply in *trans* components of the vector which have been deleted.

Representative examples of suitable expression cassettes have been described herein and include 10 synthetic Env, tat, Gag, synthetic Gag-protease, synthetic Gag-reverse transcriptase and synthetic Gag-polymerase expression cassettes, which comprise a promoter and a sequence encoding, e.g., Env, tat, or Gag-polymerase and at least one of vpr, vpu, nef or vif, 15 wherein the promoter is operably linked to Env, tat or Gag-polymerase and vpr, vpu, nef or vif. As described above, optimized Env, Gag and/or tat coding sequences may also be utilized in various combinations in the generation of packaging cell lines.

Utilizing the above-described expression cassettes, 20 a wide variety of packaging cell lines can be generated. For example, within one aspect packaging cell line are provided comprising an expression cassette that comprises a sequence encoding synthetic HIV (e.g., Gag, Env, tat, 25 Gag-polymerase, Gag-reverse transcriptase or Gag-protease) polypeptide, and a nuclear transport element, wherein the promoter is operably linked to the sequence encoding the HIV polypeptide. Within other aspects, packaging cell lines are provided comprising a promoter 30 and a sequence encoding Gag, tat, rev, or an envelope (e.g., HIV env), wherein the promoter is operably linked to the sequence encoding Gag, tat, rev, or, the envelope. Within further embodiments, the packaging cell line may comprise a sequence encoding any one or more of nef, vif,

vpu or vpr. For example, the packaging cell line may contain only nef, vif, vpu, or vpr alone, nef and vif, nef and vpu, nef and vpr, vif and vpu, vif and vpr, vpu and vpr, nef vif and vpu, nef vif and vpr, nef vpu and vpr, vvir vpu and vpr, or, all four of nef vif vpu and vpr.

In one embodiment, the expression cassette is stably integrated. Within another embodiment, the packaging cell line, upon introduction of a lentiviral vector, produces particles. Within further embodiments the promoter is inducible. Within certain preferred embodiments of the invention, the packaging cell line, upon introduction of a lentiviral vector, produces particles that are free of replication competent virus.

The synthetic cassettes containing optimized coding sequences are transfected into a selected cell line. Transfected cells are selected that (i) carry, typically, integrated, stable copies of the Gag, Pol, and Env coding sequences, and (ii) are expressing acceptable levels of these polypeptides (expression can be evaluated by methods known in the prior art, e.g., see Examples 1-4). The ability of the cell line to produce VLPs may also be verified (Examples 6, 7 and 15).

A sequence of interest is constructed into a suitable viral vector as discussed above. This defective virus is then transfected into the packaging cell line. The packaging cell line provides the viral functions necessary for producing virus-like particles into which the defective viral genome, containing the sequence of interest, are packaged. These VLPs are then isolated and can be used, for example, in gene delivery or gene therapy.

Further, such packaging cell lines can also be used to produce VLPs alone, which can, for example, be used as

adjuvants for administration with other antigens or in vaccine compositions. Also, co-expression of a selected sequence of interest encoding a polypeptide (for example, an antigen) in the packaging cell line can also result in 5 the entrapment and/or association of the selected polypeptide in/with the VLPs.

2.3 DNA IMMUNIZATION AND GENE DELIVERY

A variety of polypeptide antigens can be used in the 10 practice of the present invention. Polypeptide antigens can be included in DNA immunization constructs containing, for example, any of the synthetic expression cassettes described herein fused in-frame to a coding sequence for the polypeptide antigen, where expression of 15 the construct results in VLPs presenting the antigen of interest. Antigens can be derived from a wide variety of viruses, bacteria, fungi, plants, protozoans and other parasites. For example, the present invention will find use for stimulating an immune response against a wide 20 variety of proteins from the herpesvirus family, including proteins derived from herpes simplex virus (HSV) types 1 and 2, such as HSV-1 and HSV-2 gB, gD, gH, VP16 and VP22; antigens derived from varicella zoster virus (VZV), Epstein-Barr virus (EBV) and cytomegalovirus 25 (CMV) including CMV gB and gH; and antigens derived from other human herpesviruses such as HHV6 and HHV7. (See, e.g. Chee et al., *Cytomegaloviruses* (J.K. McDougall, ed., Springer-Verlag 1990) pp. 125-169, for a review of the protein coding content of cytomegalovirus; McGeoch et 30 al., *J. Gen. Virol.* (1988) 69:1531-1574, for a discussion of the various HSV-1 encoded proteins; U.S. Patent No. 5,171,568 for a discussion of HSV-1 and HSV-2 gB and gD proteins and the genes encoding therefore; Baer et al., *Nature* (1984) 310:207-211, for the identification of

protein coding sequences in an EBV genome; and Davison and Scott, *J. Gen. Virol.* (1986) 67:1759-1816, for a review of VZV.)

Additionally, immune responses to antigens from the hepatitis family of viruses, including hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), the delta hepatitis virus (HDV), hepatitis E virus (HEV), and hepatitis G virus, can also be stimulated using the constructs of the present invention. By way of example, the HCV genome encodes several viral proteins, including E1 (also known as E) and E2 (also known as E2/NS1), which will find use with the present invention (see, Houghton et al. *Hepatology* (1991) 14:381-388, for a discussion of HCV proteins, including E1 and E2). The δ-antigen from HDV can also be used (see, e.g., U.S. Patent No. 5,389,528, for a description of the δ-antigen).

Similarly, influenza virus is another example of a virus for which the present invention will be particularly useful. Specifically, the envelope glycoproteins HA and NA of influenza A are of particular interest for generating an immune response. Numerous HA subtypes of influenza A have been identified (Kawaoka et al., *Virology* (1990) 179:759-767; Webster et al. "Antigenic variation among type A influenza viruses," p. 127-168. In: P. Palese and D.W. Kingsbury (ed.), *Genetics of influenza viruses*. Springer-Verlag, New York).

Other antigens of particular interest to be used in the practice of the present invention include antigens and polypeptides derived therefrom from human papillomavirus (HPV), such as one or more of the various early proteins including E6 and E7; tick-borne encephalitis viruses; and HIV-1 (also known as HTLV-III, LAV, ARV, etc.), including, but not limited to, antigens such as gp120, gp41, gp160, Gag and pol from a variety of

isolates including, but not limited to, HIV_{IIIb}, HIV_{SF2}, HIV-1_{SF162}, HIV-1_{SF170}, HIV_{LAV}, HIV_{LAI}, HIV_{MN}, HIV-1_{CM235}, HIV-1_{US4}, other HIV-1 strains from diverse subtypes (e.g., subtypes, A through G, and O), HIV-2 strains and diverse 5 subtypes (e.g., HIV-2_{UC1} and HIV-2_{UC2}). See, e.g., Myers, et al., Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico; Myers, et al., *Human Retroviruses and Aids*, 1990, Los Alamos, New Mexico: Los Alamos National Laboratory.

10 Proteins derived from other viruses will also find use in the claimed methods, such as without limitation, proteins from members of the families Picornaviridae (e.g., polioviruses, etc.); Caliciviridae; Togaviridae (e.g., rubella virus, dengue virus, etc.); Flaviviridae; 15 Coronaviridae; Reoviridae; Birnaviridae; Rhabdoviridae (e.g., rabies virus, etc.); Filoviridae; Paramyxoviridae (e.g., mumps virus, measles virus, respiratory syncytial virus, etc.); Orthomyxoviridae (e.g., influenza virus types A, B and C, etc.); Bunyaviridae; Arenaviridae; 20 Retroviridae, e.g., HTLV-I; HTLV-II; HIV-1; HIV-2; simian immunodeficiency virus (SIV) among others. See, e.g. *Virology*, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991; *Virology*, 3rd Edition (Fields, BN, DM Knipe, PM 25 Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA) for a description of these and other viruses.

Particularly preferred bacterial antigens are derived from organisms that cause diphtheria, tetanus, pertussis, meningitis, and other pathogenic states, 30 including, without limitation, antigens derived from *Corynebacterium diphtheriae*, *Clostridium tetani*, *Bordetella pertusis*, *Neisseria meningitidis*, including serotypes *Meningococcus A, B, C, Y* and WI35 (*MenA, B, C, Y* and WI35), *Haemophilus influenza* type B (Hib), and

Helicobacter pylori. Examples of parasitic antigens include those derived from organisms causing malaria, tuberculosis, and Lyme disease.

Furthermore, the methods described herein provide means for treating a variety of malignant cancers. For example, the system of the present invention can be used to enhance both humoral and cell-mediated immune responses to particular proteins specific to a cancer in question, such as an activated oncogene, a fetal antigen, or an activation marker. Such tumor antigens include any of the various MAGEs (melanoma associated antigen E), including MAGE 1, 2, 3, 4, etc. (Boon, T. *Scientific American* (March 1993):82-89); any of the various tyrosinases; MART 1 (melanoma antigen recognized by T cells), mutant ras; mutant p53; p97 melanoma antigen; CEA (carcinoembryonic antigen), among others.

DNA immunization using synthetic expression cassettes of the present invention has been demonstrated to be efficacious (Examples 8 and 10-12). Animals were immunized with both the synthetic expression cassette and the wild type expression cassette. The results of the immunizations with plasmid-DNAs showed that the synthetic expression cassettes provide a clear improvement of immunogenicity relative to the native expression cassettes. Also, the second boost immunization induced a secondary immune response, for example after two to eight weeks. Further, the results of CTL assays showed increased potency of synthetic expression cassettes for induction of cytotoxic T-lymphocyte (CTL) responses by DNA immunization.

It is readily apparent that the subject invention can be used to mount an immune response to a wide variety of antigens and hence to treat or prevent a large number of diseases.

2.3.1 DELIVERY OF THE SYNTHETIC EXPRESSION CASSETTES OF THE
PRESENT INVENTION

Polynucleotide sequences coding for the above-described molecules can be obtained using recombinant methods, such as by screening cDNA and genomic libraries from cells expressing the gene, or by deriving the gene from a vector known to include the same. The sequences can be analyzed by conventional sequencing techniques. Furthermore, the desired gene can be isolated directly from cells and tissues containing the same, using standard techniques, such as phenol extraction and PCR of cDNA or genomic DNA. See, e.g., Sambrook et al., *supra*, for a description of techniques used to obtain, isolate and sequence DNA. Once the sequence is known, the gene of interest can also be produced synthetically, rather than cloned. The nucleotide sequence can be designed with the appropriate codons for the particular amino acid sequence desired. In general, one will select preferred codons for the intended host in which the sequence will be expressed. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, e.g., Edge, *Nature* (1981) 292:756; Nambair et al., *Science* (1984) 223:1299; Jay et al., *J. Biol. Chem.* (1984) 259:6311; Stemmer, W.P.C., (1995) *Gene* 164:49-53.

Next, the gene sequence encoding the desired antigen can be inserted into a vector containing a synthetic expression cassette of the present invention (e.g., see Example 1 for construction of various exemplary synthetic expression cassette). The antigen is inserted into the synthetic coding sequence such that when the combined sequence is expressed it results in the production of VLPs comprising the polypeptide and/or the antigen of

interest. Insertions can be made within the Gag coding sequence or at either end of the coding sequence (5', amino terminus of the expressed polypeptide; or 3', carboxy terminus of the expressed polypeptide -- e.g., 5 see Example 1) (Wagner, R., et al., *Arch Virol.* 127:117-137, 1992; Wagner, R., et al., *Virology* 200:162-175, 1994; Wu, X., et al., *J. Virol.* 69(6):3389-3398, 1995; Wang, C-T., et al., *Virology* 200:524-534, 1994; Chazal, N., et al., *Virology* 68(1):111-122, 1994; Griffiths, 10 J.C., et al., *J. Virol.* 67(6):3191-3198, 1993; Reicin, A.S., et al., *J. Virol.* 69(2):642-650, 1995).

Up to 50% of the coding sequences of p55Gag can be deleted without affecting the assembly to virus-like particles and expression efficiency (Borsetti, A., et al., 15 *J. Virol.* 72(11):9313-9317, 1998; Gamier, L., et al., *J. Virol.* 72(6):4667-4677, 1998; Zhang, Y., et al., *J. Virol.* 72(3):1782-1789, 1998; Wang, C., et al., *J. Virol.* 72(10):7950-7959, 1998). In one embodiment of the present invention, immunogenicity of the high level expressing 20 synthetic p55GagMod and p55GagProtMod expression cassettes can be increased by the insertion of different structural or non-structural HIV antigens, multiepitope cassettes, or cytokine sequences into deleted, mutated or truncated regions of p55GagMod sequence. In another 25 embodiment of the present invention, immunogenicity of the high level expressing synthetic Env expression cassettes can be increased by the insertion of different structural or non-structural HIV antigens, multiepitope cassettes, or cytokine sequences into deleted regions of 30 gp120Mod, gp140Mod or gp160Mod sequences. Such deletions may be generated following the teachings of the present invention and information available to one of ordinary skill in the art. One possible advantage of this approach, relative to using full-length modified Env

sequences fused to heterologous polypeptides, can be higher expression/secretion efficiency and/or higher immunogenicity of the expression product. Such deletions may be generated following the teachings of the present invention and information available to one of ordinary skill in the art. One possible advantage of this approach, relative to using full-length Env, Gag or Tat sequences fused to heterologous polypeptides, can be higher expression/secretion efficiency and/or immunogenicity of the expression product.

When sequences are added to the amino terminal end of Gag (for example, when using the synthetic p55GagMod expression cassette of the present invention), the polynucleotide can contain coding sequences at the 5' end that encode a signal for addition of a myristic moiety to the Gag-containing polypeptide (e.g., sequences that encode Met-Gly).

The ability of Gag-containing polypeptide constructs to form VLPs can be empirically determined following the teachings of the present specification.

HIV polypeptide/antigen synthetic expression cassettes include control elements operably linked to the coding sequence, which allow for the expression of the gene *in vivo* in the subject species. For example, typical promoters for mammalian cell expression include the SV40 early promoter, a CMV promoter such as the CMV immediate early promoter, the mouse mammary tumor virus LTR promoter, the adenovirus major late promoter (Ad MLP), and the herpes simplex virus promoter, among others. Other nonviral promoters, such as a promoter derived from the murine metallothionein gene, will also find use for mammalian expression. Typically, transcription termination and polyadenylation sequences will also be present, located 3' to the translation stop

codon. Preferably, a sequence for optimization of initiation of translation, located 5' to the coding sequence, is also present. Examples of transcription terminator/polyadenylation signals include those derived from SV40, as described in Sambrook et al., *supra*, as well as a bovine growth hormone terminator sequence.

Enhancer elements may also be used herein to increase expression levels of the mammalian constructs. Examples include the SV40 early gene enhancer, as described in Dijkema et al., *EMBO J.* (1985) 4:761, the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus, as described in Gorman et al., *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777 and elements derived from human CMV, as described in Boshart et al., *Cell* (1985) 41:521, such as elements included in the CMV intron A sequence.

Furthermore, plasmids can be constructed which include a chimeric antigen-coding gene sequences, encoding, e.g., multiple antigens/epitopes of interest, for example derived from a single or from more than one viral isolate.

Typically the antigen coding sequences precede or follow the synthetic coding sequences and the chimeric transcription unit will have a single open reading frame encoding both the antigen of interest and the synthetic Gag coding sequences. Alternatively, multi-cistronic cassettes (e.g., bi-cistronic cassettes) can be constructed allowing expression of multiple antigens from a single mRNA using the EMCV IRES, or the like. Lastly, antigens can be encoded on separate transcripts from independent promoters on a single plasmid or other vector.

Once complete, the constructs are used for nucleic acid immunization or the like using standard gene

delivery protocols. Methods for gene delivery are known in the art. See, e.g., U.S. Patent Nos. 5,399,346, 5,580,859, 5,589,466. Genes can be delivered either directly to the vertebrate subject or, alternatively, 5 delivered *ex vivo*, to cells derived from the subject and the cells reimplanted in the subject.

A number of viral based systems have been developed for gene transfer into mammalian cells. For example, retroviruses provide a convenient platform for gene 10 delivery systems. Selected sequences can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of retroviral 15 systems have been described (U.S. Patent No. 5,219,740; Miller and Rosman, *BioTechniques* (1989) 7:980-990; Miller, A.D., *Human Gene Therapy* (1990) 1:5-14; Scarpa et al., *Virology* (1991) 180:849-852; Burns et al., *Proc. Natl. Acad. Sci. USA* (1993) 90:8033-8037; and Boris- 20 Lawrie and Temin, *Cur. Opin. Genet. Develop.* (1993) 3:102-109.

A number of adenovirus vectors have also been described. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus 25 minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham, *J. Virol.* (1986) 57:267-274; Bett et al., *J. Virol.* (1993) 67:5911-5921; Mittereder et al., *Human Gene Therapy* (1994) 5:717-729; Seth et al., *J. Virol.* (1994) 68:933-940; Barr et al., 30 *Gene Therapy* (1994) 1:51-58; Berkner, K.L. *BioTechniques* (1988) 6:616-629; and Rich et al., *Human Gene Therapy* (1993) 4:461-476).

Additionally, various adeno-associated virus (AAV) vector systems have been developed for gene delivery.

AAV vectors can be readily constructed using techniques well known in the art. See, e.g., U.S. Patent Nos. 5,173,414 and 5,139,941; International Publication Nos. WO 92/01070 (published 23 January 1992) and WO 93/03769 (published 4 March 1993); Lebkowski et al., *Molec. Cell. Biol.* (1988) 8:3988-3996; Vincent et al., *Vaccines* 90 (1990) (Cold Spring Harbor Laboratory Press); Carter, B.J. *Current Opinion in Biotechnology* (1992) 3:533-539; Muzychka, N. *Current Topics in Microbiol. and Immunol.* (1992) 158:97-129; Kotin, R.M. *Human Gene Therapy* (1994) 5:793-801; Shelling and Smith, *Gene Therapy* (1994) 1:165-169; and Zhou et al., *J. Exp. Med.* (1994) 179:1867-1875.

Another vector system useful for delivering the polynucleotides of the present invention is the enterically administered recombinant poxvirus vaccines described by Small, Jr., P.A., et al. (U.S. Patent No. 5,676,950, issued October 14, 1997).

Additional viral vectors which will find use for delivering the nucleic acid molecules encoding the antigens of interest include those derived from the pox family of viruses, including vaccinia virus and avian poxvirus. By way of example, vaccinia virus recombinants expressing the genes can be constructed as follows. The DNA encoding the particular synthetic Gag/antigen coding sequence is first inserted into an appropriate vector so that it is adjacent to a vaccinia promoter and flanking vaccinia DNA sequences, such as the sequence encoding thymidine kinase (TK). This vector is then used to transfect cells which are simultaneously infected with vaccinia. Homologous recombination serves to insert the vaccinia promoter plus the gene encoding the coding sequences of interest into the viral genome. The resulting TK recombinant can be selected by culturing the

cells in the presence of 5-bromodeoxyuridine and picking viral plaques resistant thereto.

Alternatively, avipoxviruses, such as the fowlpox and canarypox viruses, can also be used to deliver the genes. Recombinant avipox viruses, expressing immunogens from mammalian pathogens, are known to confer protective immunity when administered to non-avian species. The use of an avipox vector is particularly desirable in human and other mammalian species since members of the avipox genus can only productively replicate in susceptible avian species and therefore are not infective in mammalian cells. Methods for producing recombinant avipoxviruses are known in the art and employ genetic recombination, as described above with respect to the production of vaccinia viruses. See, e.g., WO 91/12882; WO 89/03429; and WO 92/03545.

Molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al., *J. Biol. Chem.* (1993) 268:6866-6869 and Wagner et al., *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103, can also be used for gene delivery.

Members of the Alphavirus genus, such as, but not limited to, vectors derived from the Sindbis, Semliki Forest, and Venezuelan Equine Encephalitis viruses, will also find use as viral vectors for delivering the polynucleotides of the present invention (for example, a synthetic Gag- or Env-polypeptide encoding expression cassette as described in Example 14 below). For a description of Sindbis-virus derived vectors useful for the practice of the instant methods, see, Dubensky et al., *J. Virol.* (1996) 70:508-519; and International Publication Nos. WO 95/07995 and WO 96/17072; as well as, Dubensky, Jr., T.W., et al., U.S. Patent No. 5,843,723,

issued December 1, 1998, and Dubensky, Jr., T.W., U.S. Patent No. 5,789,245, issued August 4, 1998.

A vaccinia based infection/transfection system can be conveniently used to provide for inducible, transient expression of the coding sequences of interest (for example, a synthetic Gag/HCV-core expression cassette) in a host cell. In this system, cells are first infected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the polynucleotide of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into protein by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation products. See, e.g., Elroy-Stein and Moss, *Proc. Natl. Acad. Sci. USA* (1990) 87:6743-6747; Fuerst et al., *Proc. Natl. Acad. Sci. USA* (1986) 83:8122-8126.

As an alternative approach to infection with vaccinia or avipox virus recombinants, or to the delivery of genes using other viral vectors, an amplification system can be used that will lead to high level expression following introduction into host cells. Specifically, a T7 RNA polymerase promoter preceding the coding region for T7 RNA polymerase can be engineered. Translation of RNA derived from this template will generate T7 RNA polymerase which in turn will transcribe more template. Concomitantly, there will be a cDNA whose expression is under the control of the T7 promoter. Thus, some of the T7 RNA polymerase generated from

translation of the amplification template RNA will lead to transcription of the desired gene. Because some T7 RNA polymerase is required to initiate the amplification, T7 RNA polymerase can be introduced into cells along with 5 the template(s) to prime the transcription reaction. The polymerase can be introduced as a protein or on a plasmid encoding the RNA polymerase. For a further discussion of T7 systems and their use for transforming cells, see, e.g., International Publication No. WO 94/26911; Studier 10 and Moffatt, *J. Mol. Biol.* (1986) 189:113-130; Deng and Wolff, *Gene* (1994) 143:245-249; Gao et al., *Biochem. Biophys. Res. Commun.* (1994) 200:1201-1206; Gao and Huang, *Nuc. Acids Res.* (1993) 21:2867-2872; Chen et al., *Nuc. Acids Res.* (1994) 22:2114-2120; and U.S. Patent No. 15 5,135,855.

The synthetic expression cassette of interest can also be delivered without a viral vector. For example, the synthetic expression cassette can be packaged as DNA or RNA in liposomes prior to delivery to the subject or 20 to cells derived therefrom. Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed DNA to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or 25 more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight, *Biochim. Biophys. Acta.* (1991) 1097:1-17; Straubinger et al., in *Methods of Enzymology* (1983), Vol. 101, pp. 512-527.

30 Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations, with cationic liposomes particularly preferred. Cationic liposomes have been shown to mediate intracellular

delivery of plasmid DNA (Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-7416); mRNA (Malone et al., *Proc. Natl. Acad. Sci. USA* (1989) 86:6077-6081); and purified transcription factors (Debs et al., *J. Biol. Chem.* (1990) 265:10189-10192), in functional form.

5 Cationic liposomes are readily available. For example, N[1-2,3-dioleyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. 10 (See, also, Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-7416). Other commercially available lipids include (DDAB/DOPE) and DOTAP/DOPE (Boerhinger). Other cationic liposomes can be prepared from readily available materials using techniques well known in the 15 art. See, e.g., Szoka et al., *Proc. Natl. Acad. Sci. USA* (1978) 75:4194-4198; PCT Publication No. WO 90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

20 Similarly, anionic and neutral liposomes are readily available, such as, from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl 25 glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

30 The liposomes can comprise multilammellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See, e.g., Straubinger et al., in METHODS OF

IMMUNOLOGY (1983), Vol. 101, pp. 512-527; Szoka et al.,
Proc. Natl. Acad. Sci. USA (1978) 75:4194-4198;
Papahadjopoulos et al., Biochim. Biophys. Acta (1975)
394:483; Wilson et al., Cell (1979) 17:77); Deamer and
5 Bangham, Biochim. Biophys. Acta (1976) 443:629; Ostro et
al., Biochem. Biophys. Res. Commun. (1977) 76:836; Fraley
et al., Proc. Natl. Acad. Sci. USA (1979) 76:3348; Enoch
and Strittmatter, Proc. Natl. Acad. Sci. USA (1979)
76:145); Fraley et al., J. Biol. Chem. (1980) 255:10431;
10 Szoka and Papahadjopoulos, Proc. Natl. Acad. Sci. USA
(1978) 75:145; and Schaefer-Ridder et al., Science (1982)
215:166.

The DNA and/or protein antigen(s) can also be delivered in cochleate lipid compositions similar to
15 those described by Papahadjopoulos et al., Biochem.
Biophys. Acta. (1975) 394:483-491. See, also, U.S.
Patent Nos. 4,663,161 and 4,871,488.

The synthetic expression cassette of interest (e.g., any of the synthetic expression cassettes described in
20 Example 1) may also be encapsulated, adsorbed to, or associated with, particulate carriers. Such carriers present multiple copies of a selected antigen to the immune system and promote migration, trapping and retention of antigens in local lymph nodes. The
25 particles can be taken up by profession antigen presenting cells such as macrophages and dendritic cells, and/or can enhance antigen presentation through other mechanisms such as stimulation of cytokine release. Examples of particulate carriers include those derived
30 from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-glycolides), known as PLG. See, e.g., Jeffery et al., Pharm. Res. (1993) 10:362-368; McGee JP,

et al., *J Microencapsul.* 14(2):197-210, 1997; O'Hagan DT, et al., *Vaccine* 11(2):149-54, 1993.

Furthermore, other particulate systems and polymers can be used for the *in vivo* or *ex vivo* delivery of the gene of interest. For example, polymers such as polylysine, polyarginine, polyornithine, spermine, spermidine, as well as conjugates of these molecules, are useful for transferring a nucleic acid of interest. Similarly, DEAE dextran-mediated transfection, calcium phosphate precipitation or precipitation using other insoluble inorganic salts, such as strontium phosphate, aluminum silicates including bentonite and kaolin, chromic oxide, magnesium silicate, talc, and the like, will find use with the present methods. See, e.g., 10 Felgner, P.L., *Advanced Drug Delivery Reviews* (1990) 5:163-187, for a review of delivery systems useful for gene transfer. Peptoids (Zuckerman, R.N., et al., U.S. Patent No. 5,831,005, issued November 3, 1998) may also be used for delivery of a construct of the present 15 invention.

Additionally, biolistic delivery systems employing particulate carriers such as gold and tungsten, are especially useful for delivering synthetic expression cassettes of the present invention. The particles are 20 coated with the synthetic expression cassette(s) to be delivered and accelerated to high velocity, generally under a reduced atmosphere, using a gun powder discharge from a "gene gun." For a description of such techniques, and apparatuses useful therefore, see, e.g., U.S. Patent 25 Nos. 4,945,050; 5,036,006; 5,100,792; 5,179,022; 30 5,371,015; and 5,478,744. Also, needle-less injection systems can be used (Davis, H.L., et al, *Vaccine* 12:1503-1509, 1994; Bioject, Inc., Portland, OR).

Recombinant vectors carrying a synthetic expression cassette of the present invention are formulated into compositions for delivery to the vertebrate subject. These compositions may either be prophylactic (to prevent infection) or therapeutic (to treat disease after infection). The compositions will comprise a "therapeutically effective amount" of the gene of interest such that an amount of the antigen can be produced *in vivo* so that an immune response is generated in the individual to which it is administered. The exact amount necessary will vary depending on the subject being treated; the age and general condition of the subject to be treated; the capacity of the subject's immune system to synthesize antibodies; the degree of protection desired; the severity of the condition being treated; the particular antigen selected and its mode of administration, among other factors. An appropriate effective amount can be readily determined by one of skill in the art. Thus, a "therapeutically effective amount" will fall in a relatively broad range that can be determined through routine trials.

The compositions will generally include one or more "pharmaceutically acceptable excipients or vehicles" such as water, saline, glycerol, polyethyleneglycol, hyaluronic acid, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, surfactants and the like, may be present in such vehicles. Certain facilitators of immunogenicity or of nucleic acid uptake and/or expression can also be included in the compositions or coadministered, such as, but not limited to, bupivacaine, cardiotoxin and sucrose.

Once formulated, the compositions of the invention can be administered directly to the subject (e.g., as

described above) or, alternatively, delivered *ex vivo*, to cells derived from the subject, using methods such as those described above. For example, methods for the *ex vivo* delivery and reimplantation of transformed cells 5 into a subject are known in the art and can include, e.g., dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, lipofectamine and LT-1 mediated transfection, protoplast fusion, electroporation, encapsulation of the 10 polynucleotide(s) (with or without the corresponding antigen) in liposomes, and direct microinjection of the DNA into nuclei.

Direct delivery of synthetic expression cassette compositions *in vivo* will generally be accomplished with 15 or without viral vectors, as described above, by injection using either a conventional syringe, needless devices such as Bioject® or a gene gun, such as the Accell® gene delivery system (PowderJect Technologies, Inc., Oxford, England). The constructs can be delivered 20 (e.g., injected) either subcutaneously, epidermally, intradermally, intramuscularly, intravenous, intramucosally (such as nasally, rectally and vaginally), intraperitoneally or orally. Delivery of DNA into cells of the epidermis is particularly preferred as this mode 25 of administration provides access to skin-associated lymphoid cells and provides for a transient presence of DNA in the recipient. Other modes of administration include oral ingestion and pulmonary administration, suppositories, needle-less injection, transcutaneous and 30 transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule.

**2.3.2 EX VIVO DELIVERY OF THE SYNTHETIC EXPRESSION
CASSETTES OF THE PRESENT INVENTION**

In one embodiment, T cells, and related cell types (including but not limited to antigen presenting cells, such as, macrophage, monocytes, lymphoid cells, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof), can be used for ex vivo delivery of the synthetic expression cassettes of the present invention. T cells can be isolated from peripheral blood lymphocytes (PBLs) by a variety of procedures known to those skilled in the art. For example, T cell populations can be "enriched" from a population of PBLs through the removal of accessory and B cells. In particular, T cell enrichment can be accomplished by the elimination of non-T cells using anti-MHC class II monoclonal antibodies. Similarly, other antibodies can be used to deplete specific populations of non-T cells. For example, anti-Ig antibody molecules can be used to deplete B cells and anti-MacI antibody molecules can be used to deplete macrophages.

T cells can be further fractionated into a number of different subpopulations by techniques known to those skilled in the art. Two major subpopulations can be isolated based on their differential expression of the cell surface markers CD4 and CD8. For example, following the enrichment of T cells as described above, CD4⁺ cells can be enriched using antibodies specific for CD4 (see Coligan et al., *supra*). The antibodies may be coupled to a solid support such as magnetic beads. Conversely, CD8⁺ cells can be enriched through the use of antibodies specific for CD4 (to remove CD4⁺ cells), or can be isolated by the use of CD8 antibodies coupled to a solid support. CD4

lymphocytes from HIV-1 infected patients can be expanded ex vivo, before or after transduction as described by Wilson et. al. (1995) *J. Infect. Dis.* 172:88.

5 Following purification of T cells, a variety of methods of genetic modification known to those skilled in the art can be performed using non-viral or viral-based gene transfer vectors constructed as described herein. For example, one such approach involves transduction of
10 the purified T cell population with vector-containing supernatant of cultures derived from vector producing cells. A second approach involves co-cultivation of an irradiated monolayer of vector-producing cells with the purified T cells. A third approach involves a similar
15 co-cultivation approach; however, the purified T cells are pre-stimulated with various cytokines and cultured 48 hours prior to the co-cultivation with the irradiated vector producing cells. Pre-stimulation prior to such transduction increases effective gene transfer (Nolta et
20 al. (1992) *Exp. Hematol.* 20:1065). Stimulation of these cultures to proliferate also provides increased cell populations for re-infusion into the patient. Subsequent to co-cultivation, T cells are collected from the vector producing cell monolayer, expanded, and frozen in liquid
25 nitrogen.

Gene transfer vectors, containing one or more synthetic expression cassette of the present invention (associated with appropriate control elements for delivery to the isolated T cells) can be assembled using
30 known methods.

Selectable markers can also be used in the construction of gene transfer vectors. For example, a marker can be used which imparts to a mammalian cell transduced with the gene transfer vector resistance to a

cytotoxic agent. The cytotoxic agent can be, but is not limited to, neomycin, aminoglycoside, tetracycline, chloramphenicol, sulfonamide, actinomycin, netropsin, distamycin A, anthracycline, or pyrazinamide. For example, neomycin phosphotransferase II imparts resistance to the neomycin analogue geneticin (G418).

The T cells can also be maintained in a medium containing at least one type of growth factor prior to being selected. A variety of growth factors are known in the art which sustain the growth of a particular cell type. Examples of such growth factors are cytokine mitogens such as rIL-2, IL-10, IL-12, and IL-15, which promote growth and activation of lymphocytes. Certain types of cells are stimulated by other growth factors such as hormones, including human chorionic gonadotropin (hCG) and human growth hormone. The selection of an appropriate growth factor for a particular cell population is readily accomplished by one of skill in the art.

For example, white blood cells such as differentiated progenitor and stem cells are stimulated by a variety of growth factors. More particularly, IL-3, IL-4, IL-5, IL-6, IL-9, GM-CSF, M-CSF, and G-CSF, produced by activated T_h and activated macrophages, stimulate myeloid stem cells, which then differentiate into pluripotent stem cells, granulocyte-monocyte progenitors, eosinophil progenitors, basophil progenitors, megakaryocytes, and erythroid progenitors. Differentiation is modulated by growth factors such as GM-CSF, IL-3, IL-6, IL-11, and EPO.

Pluripotent stem cells then differentiate into lymphoid stem cells, bone marrow stromal cells, T cell progenitors, B cell progenitors, thymocytes, T_h Cells, T_c cells, and B cells. This differentiation is modulated by

growth factors such as IL-3, IL-4, IL-6, IL-7, GM-CSF, M-CSF, G-CSF, IL-2, and IL-5.

5 Granulocyte-monocyte progenitors differentiate to monocytes, macrophages, and neutrophils. Such differentiation is modulated by the growth factors GM-CSF, M-CSF, and IL-8. Eosinophil progenitors differentiate into eosinophils. This process is modulated by GM-CSF and IL-5.

10 The differentiation of basophil progenitors into mast cells and basophils is modulated by GM-CSF, IL-4, and IL-9. Megakaryocytes produce platelets in response to GM-CSF, EPO, and IL-6. Erythroid progenitor cells differentiate into red blood cells in response to EPO.

15 Thus, during activation by the CD3-binding agent, T cells can also be contacted with a mitogen, for example a cytokine such as IL-2. In particularly preferred embodiments, the IL-2 is added to the population of T cells at a concentration of about 50 to 100 µg/ml. Activation with the CD3-binding agent can be carried out
20 for 2 to 4 days.

Once suitably activated, the T cells are genetically modified by contacting the same with a suitable gene transfer vector under conditions that allow for transfection of the vectors into the T cells. Genetic
25 modification is carried out when the cell density of the T cell population is between about 0.1×10^6 and 5×10^6 , preferably between about 0.5×10^6 and 2×10^6 . A number of suitable viral and nonviral-based gene transfer vectors have been described for use herein.

30 After transduction, transduced cells are selected away from non-transduced cells using known techniques. For example, if the gene transfer vector used in the transduction includes a selectable marker which confers resistance to a cytotoxic agent, the cells can be

contacted with the appropriate cytotoxic agent, whereby non-transduced cells can be negatively selected away from the transduced cells. If the selectable marker is a cell surface marker, the cells can be contacted with a binding agent specific for the particular cell surface marker, whereby the transduced cells can be positively selected away from the population. The selection step can also entail fluorescence-activated cell sorting (FACS) techniques, such as where FACS is used to select cells from the population containing a particular surface marker, or the selection step can entail the use of magnetically responsive particles as retrievable supports for target cell capture and/or background removal.

More particularly, positive selection of the transduced cells can be performed using a FACS cell sorter (e.g. a FACSVantage™ Cell Sorter, Becton Dickinson Immunocytometry Systems, San Jose, CA) to sort and collect transduced cells expressing a selectable cell surface marker. Following transduction, the cells are stained with fluorescent-labeled antibody molecules directed against the particular cell surface marker. The amount of bound antibody on each cell can be measured by passing droplets containing the cells through the cell sorter. By imparting an electromagnetic charge to droplets containing the stained cells, the transduced cells can be separated from other cells. The positively selected cells are then harvested in sterile collection vessels. These cell sorting procedures are described in detail, for example, in the FACSVantage™ Training Manual, with particular reference to sections 3-11 to 3-28 and 10-1 to 10-17.

Positive selection of the transduced cells can also be performed using magnetic separation of cells based on expression of a particular cell surface marker. In such

separation techniques, cells to be positively selected are first contacted with specific binding agent (e.g., an antibody or reagent the interacts specifically with the cell surface marker). The cells are then contacted with 5 retrievable particles (e.g., magnetically responsive particles) which are coupled with a reagent that binds the specific binding agent (that has bound to the positive cells). The cell-binding agent-particle complex can then be physically separated from non-labeled cells, 10 for example using a magnetic field. When using magnetically responsive particles, the labeled cells can be retained in a container using a magnetic filed while the negative cells are removed. These and similar separation procedures are known to those of ordinary 15 skill in the art.

Expression of the vector in the selected transduced cells can be assessed by a number of assays known to those skilled in the art. For example, Western blot or Northern analysis can be employed depending on the nature 20 of the inserted nucleotide sequence of interest. Once expression has been established and the transformed T cells have been tested for the presence of the selected synthetic expression cassette, they are ready for infusion into a patient via the peripheral blood stream.

25 The invention includes a kit for genetic modification of an *ex vivo* population of primary mammalian cells. The kit typically contains a gene transfer vector coding for at least one selectable marker and at least one synthetic expression cassette contained 30 in one or more containers, ancillary reagents or hardware, and instructions for use of the kit.

EXPERIMENTAL

Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not 5 intended to limit the scope of the present invention in any way.

Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, 10 of course, be allowed for.

Example 1**Generation of Synthetic Gag and Env Expression Cassettes**

15 A. Modification of HIV-1 Gag, Gag-protease, Gag-reverse transcriptase and Gag-polymerase Nucleic Acid Coding Sequences

The Gag (SEQ ID NO:1), Gag-protease (SEQ ID NO:2), Gag-polymerase (SEQ ID NO:3), and Gag-reverse 20 transcriptase (SEQ ID NO:77) coding sequences were selected from the HIV-1SF2 strain (Sanchez-Pescador, R., et al., *Science* 227(4686): 484-492, 1985; Luciw, P.A., et al. U.S. Patent No. 5,156,949, issued October 20, 1992; Luciw, P.A., et al., U.S. Patent No. 5,688,688, November 25 18, 1997). These sequences were manipulated to maximize expression of their gene products.

First, the HIV-1 codon usage pattern was modified so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human 30 genes. The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a high AU content in the RNA and in a decreased translation ability and instability of the

mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The Gag-encoding sequences were modified to be comparable to codon usage found in highly expressed human genes.

5 Figure 11 presents a comparison of the percent A-T content for the cDNAs of stable versus unstable RNAs (comparison window size = 50). Human IFN γ mRNA is known to (i) be unstable, (ii) have a short half-life, and (iii) have a high A-U content. Human GAPDH
10 (glyceraldehyde-3-phosphate dehydrogenase) mRNA is known to (i) be a stable RNA, and (i) have a low A-U content. In Figure 11, the percent A-T content of these two sequences are compared to the percent A-T content of native HIV-1SF2 Gag cDNA and to the synthetic Gag cDNA
15 sequence of the present invention. The top two panels of the figure show the percent A-T content over the length of the sequences for IFN γ and native Gag. The bottom two panels of the figure show the percent A-T content over the length of the sequences for GAPDH and the synthetic
20 Gag. Experiments performed in support of the present invention showed that the synthetic Gag sequences were capable of higher level of protein production (see the Examples) than the native Gag sequences. The data in Figure 11 suggest that one reason for this increased
25 production may be increased stability of the mRNA corresponding to the synthetic Gag coding sequences versus the mRNA corresponding to the native Gag coding sequences.

Second, there are inhibitory (or instability) elements (INS) located within the coding sequences of the Gag and Gag-protease coding sequences (Schneider R, et al., J Virol. 71(7):4892-4903, 1997). RRE is a secondary RNA structure that interacts with the HIV encoded Rev-protein to overcome the expression down-regulating

effects of the INS. To overcome the requirement for post-transcriptional activating mechanisms of RRE and Rev, and to enhance independent expression of the Gag polypeptide, the INS were inactivated by introducing 5 multiple point mutations that did not alter the reading frame of the encoded proteins. Figure 1 shows the original SF2 Gag sequence, the location of the INS sequences, and the modifications made to the INS sequences to reduce their effects.

10 For the Gag-protease sequence (wild type, SEQ ID NO:2; synthetic, SEQ ID NOS:5, 78 and 79), the changes in codon usage were restricted to the regions up to the -1 frameshift and starting again at the end of the Gag reading frame (Figure 2; the region indicated in lower 15 case letters in Figure 2 is the unmodified region). Further, inhibitory (or instability) elements (INS) located within the coding sequences of the Gag-protease polypeptide coding sequence were altered as well (indicated in Figure 2). The synthetic coding sequences 20 were assembled by the Midland Certified Reagent Company (Midland, Texas).

Modification of the Gag-polymerase sequences (wild type, SEQ ID NO:3; synthetic, SEQ ID NO:6) and Gag-reverse transcriptase sequences (SEQ ID NOS:80 through 25 84) include similar modifications as described for Gag-protease in order to preserve the frameshift region. Locations of the inactivation sites and changes to the sequence to alter the inactivation sites are presented in Figure 12 for the native HIV-1_{SF2} Gag-polymerase sequence.

30 In one embodiment of the invention, the full length polymerase coding region of the Gag-polymerase sequence is included with the synthetic Gag sequences in order to increase the number of epitopes for virus-like particles expressed by the synthetic, optimized Gag expression

cassette. Because synthetic HIV-1 Gag-polymerase expresses the potentially deleterious functional enzymes reverse transcriptase (RT) and integrase (INT) (in addition to the structural proteins and protease), it is important to inactivate RT and INT functions. Several in-frame deletions in the RT and INT reading frame can be made to achieve catalytic nonfunctional enzymes with respect to their RT and INT activity. {Jay. A. Levy (Editor) (1995) *The Retroviridae*, Plenum Press, New York.

10 ISBN 0-306-45033X. Pages 215-20; Grimison, B. and Laurence, J. (1995), *Journal Of Acquired Immune Deficiency Syndromes and Human Retrovirology* 9(1):58-68; Wakefield, J. K., et al., (1992) *Journal Of Virology* 66(11):6806-6812; Esnouf, R., et al., (1995) *Nature Structural Biology* 2(4):303-308; Maignan, S., et al., (1998) *Journal Of Molecular Biology* 282(2):359-368; Katz, R. A. and Skalka, A. M. (1994) *Annual Review Of Biochemistry* 73 (1994); Jacobo-Molina, A., et al., (1993) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 90(13):6320-6324; Hickman, A. B., et al., (1994) *Journal Of Biological Chemistry* 269(46):29279-29287; Goldgur, Y., et al., (1998) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 95(16):9150-9154; Goette, M., et al., (1998) *Journal Of Biological Chemistry* 273(17):10139-10146; Gorton, J. L., et al., (1998) *Journal of Virology* 72(6):5046-5055; Engelman, A., et al., (1997) *Journal Of Virology* 71(5):3507-3514; Dyda, F., et al., *Science* 266(5193):1981-1986; Davies, J. F., et al., (1991) *Science* 252(5002):88-95; Bujacz, G., et al., (1996) *Febs Letters* 398(2-3):175-178; Beard, W. A., et al., (1996) *Journal Of Biological Chemistry* 271(21):12213-12220; Kohlstaedt, L. A., et al., (1992)

Science 256(5065):1783-1790; Krug, M. S. and Berger, S. L. (1991) Biochemistry 30(44):10614-10623; Mazumder, A., et al., (1996) Molecular Pharmacology 49(4):621-628; Palaniappan, C., et al., (1997) Journal Of Biological Chemistry 272(17):11157-11164; Rodgers, D. W., et al., (1995) Proceedings Of the National Academy Of Sciences Of the United States Of America 92(4):1222-1226; Sheng, N. and Dennis, D. (1993) Biochemistry 32(18):4938-4942; Spence, R. A., et al., (1995) Science 267(5200):988-993. }

10 Furthermore selected B- and/or T-cell epitopes can be added to the Gag-polymerase constructs within the deletions of the RT- and INT-coding sequence to replace and augment any epitopes deleted by the functional modifications of RT and INT. Alternately, selected B-
15 and T-cell epitopes (including CTL epitopes) from RT and INT can be included in a minimal VLP formed by expression of the synthetic Gag or synthetic GagProt cassette, described above. (For descriptions of known HIV B- and T-cell epitopes see, HIV Molecular Immunology Database CTL
20 Search Interface; Los Alamos Sequence Compendia, 1987-1997; Internet address: <http://hiv-web.lanl.gov/immunology/index.html>.)

The resulting modified coding sequences are presented as a synthetic Gag expression cassette (SEQ ID NO:4), a synthetic Gag-protease expression cassette (SEQ ID NOs:5, 78 and 79), and a synthetic Gag-polymerase expression cassette (SEQ ID NO:6). Synthetic expression cassettes containing codon modifications in the reverse transcriptase region are shown in SEQ ID NOs:80 through 30 84. An alignment of selected sequences is presented in Figure 7. A common region (Gag-common; SEQ ID NO:9) extends from position 1 to position 1262.

The synthetic DNA fragments for Gag and Gag-protease were cloned into the following expression vectors:

pCMVKm2, for transient expression assays and DNA immunization studies, the pCMVKm2 vector was derived from pCMV6a (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986) and comprises a kanamycin selectable marker, a 5 ColE1 origin of replication, a CMV promoter enhancer and Intron A, followed by an insertion site for the synthetic sequences described below followed by a polyadenylation signal derived from bovine growth hormone -- the pCMVKm2 vector differs from the pCMV-link vector only in that a 10 polylinker site was inserted into pCMVKm2 to generate pCMV-link (Figure 14, polylinker at positions 1646 to 1697); pESN2dhfr (Figure 13A) and pCMVPLEdhfr (also known as pCMVIII as shown in Figure 13B), for expression in Chinese Hamster Ovary (CHO) cells; and, pAcC13, a shuttle 15 vector for use in the Baculovirus expression system (pAcC13, was derived from pAcC12 which was described by Munemitsu S., et al., *Mol Cell Biol.* 10(11):5977-5982, 1990).

A restriction map for vector pCMV-link is presented 20 in Figure 14. In the figure, the CMV promoter (CMV IE ENH/PRO), bovine growth hormone terminator (BGH pA), kanamycin selectable marker (kan), and a ColE1 origin of replication (ColE1 ori) are indicated. A polycloning site is also indicated in the figure following the CMV 25 promoter sequences.

A restriction map for vector pESN2dhfr is presented in Figure 13A. In the figure, the CMV promoter (pCMV, hCMVIE), bovine growth hormone terminator (BGH_pA), SV40 30 origin of replication (SV40ori), neomycin selectable marker (Neo), SV40 polyA (SV40pA), Adenovirus 2 late promoter (Ad2VLP), and the murine dhfr gene (*mu* dhfr) are indicated. A polycloning site is also indicated in the figure following the CMV promoter sequences.

Briefly, construction of pCMVPLedhfr (pCMVIII) was as follows. To construct a DHFR cassette, the EMCV IRES (internal ribosome entry site) leader was PCR-amplified from pCite-4a+ (Novagen, Inc., Milwaukee, WI) and inserted into pET-23d (Novagen, Inc., Milwaukee, WI) as an Xba-Nco fragment to give pET-EMCV. The dhfr gene was PCR-amplified from pESN2dhfr to give a product with a Gly-Gly-Gly-Ser spacer in place of the translation stop codon and inserted as an Nco-BamH1 fragment to give pET-E-DHFR. Next, the attenuated neo gene was PCR amplified from a pSV2Neo (Clontech, Palo Alto, CA) derivative and inserted into the unique BamH1 site of pET-E-DHFR to give pET-E-DHFR/Neo_(m2). Then, the bovine growth hormone terminator from pCDNA3 (Invitrogen, Inc., Carlsbad, CA) was inserted downstream of the neo gene to give pET-E-DHFR/Neo_(m2)BGHt. The EMCV-dhfr/neo selectable marker cassette fragment was prepared by cleavage of pET-E-DHFR/Neo_(m2)BGHt. The CMV enhancer/promoter plus Intron A was transferred from pCMV6a (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986) as a HindIII-SalI fragment into pUC19 (New England Biolabs, Inc., Beverly, MA). The vector backbone of pUC19 was deleted from the NdeI to the SphI sites. The above described DHFR cassette was added to the construct such that the EMCV IRES followed the CMV promoter to produce the final construct. The vector also contained an amp^r gene and an SV40 origin of replication.

Selected pCMVKm2 vectors containing the synthetic expression cassettes have been designated as follows: pCMVKm2.GagMod.SF2, pCMVKm2.GagprotMod.SF2, and pCMVKm2.GagpolMod.SF2, pCMVKm2.GagprotMod.SF2.GP1 (SEQ ID NO:78) and pCMVKm2.GagprotMod.SF2.GP2 (SEQ ID NO:79). Other exemplary Gag-encoding expressing cassettes are shown in the Figures and as Sequence Listings.

B. Modification of HIV-1 Gag/Hepatitis C Core Chimeric Protein Nucleic Acid Coding Sequences Generation of Synthetic Expression Cassettes

To facilitate the ligation of the Gag and HCV core coding sequences, PCR amplification was employed. The synthetic p55Gag expression cassette was used as a PCR template with the following primers: GAG5 (SEQ ID NO:11) and P55-SAL3 (SEQ ID NO:12). The PCR amplification was conducted at 55°C for 25 cycles using Stratagene's Pfu polymerase. The resulting PCR product was rendered free of nucleotides and primers using the Promega PCR clean-up kit and then subjected to EcoRI and SalI digestions. For HCV core coding sequences, the following primers were used with an HCV template (Houghton, M., et al., U.S. Patent No. 5,714,596, issued February 3, 1998; Houghton, M., et al., U.S. Patent No. 5,712,088, issued January 27, 1998; Houghton, M., et al., U.S. Patent No. 5,683,864, issued November 4, 1997; Weiner, A.J., et al., U.S. Patent No. 5,728,520, issued March 17, 1998; Weiner, A.J., et al., U.S. Patent No. 5,766,845, issued June 16, 1998; Weiner, A.J., et al., U.S. Patent No. 5,670,152, issued September 23, 1997): CORESAL 5 (SEQ ID NO:13) and 173CORE (SEQ ID NO:14) using the conditions outlined above. The purified product was digested with SalI and BamHI restriction enzymes. The digested Gag and HCV core PCR products were ligated into the pCMVKm2 vector digested with EcoRI and BamHI. Ligation of the PCR products at the SalI site resulted in a direct fusion of the final amino acid of p55Gag to the second amino acid of HCV core, serine. Amino acid 173 of core is a serine and is followed immediately by a TAG termination codon. The sequence of the fusion clone was confirmed. The pCMVKm2 vector containing the synthetic expression

cassette was designated as pCMVKm2.GagModHCVcore.

The EcoRI-BamHI fragment of p55Gag-core 173 was also cloned into EcoRI-BamHI-digested pAcC13 for baculovirus expression. Western blots confirmed expression and 5 sucrose gradient sedimentation along with electron microscopy confirmed particle formation. To generate the above clone but containing the synthetic Gag sequences (instead of wild-type), the following steps were performed: pCMVKm2-modified p55Gag was used as template 10 for PCR amplification with MS65 (SEQ ID NO:15) and MS66 (SEQ ID NO:16) primers. The region amplified corresponds to the BspHI and SalI sites at the C-terminus 15 of synthetic Gag sequence. The amplification product was digested with BspHI and SalI and ligated to SalI/BamHI digested pCMV-link along with the SalI/BspHI fragment from 20 pCMV-Km-p55modGag , representing the amino terminal end of modified Gag, and the SalI/BamHI fragment from pCMV-p55Gag-core173. Thereafter, a T4-blunted-SalI partial/BamHI fragment was ligated into pAcC4-SmaI/BamHI to generate pAcC4-p55GagMod-core173 (containing the 25 synthetic sequence presented as SEQ ID NO:7).

C. Defining of the Major Homology Region (MHR) of HIV-1 p55Gag

25 The Major Homology Region (MHR) of HIV-1 p55 (Gag) is located in the p24-CA sequence of Gag. It is a conserved stretch of 20 amino acids (SEQ ID NO:19). The position in the wild type HIV-1_{SP2} Gag protein is from aa 286-305 and spans a region from nucleotides 856-915 in 30 the native HIV-1_{SP2} Gag DNA-sequence. The position in the synthetic Gag protein is from aa 288-307 and spans a region from nucleotides 862-921 for the synthetic Gag DNA-sequence. The nucleotide sequence for the MHR in the synthetic

GagMod.SF2 is presented as SEQ ID NO:20. Mutations or deletions in the amino acid sequence of the MHR can severely impair particle production (Borsetti, A., et al., *J. Virol.* 72(11):9313-9317, 1998; Mammano, F., et al., *J Virol* 68(8):4927-4936, 1994).

Percent identity to the MHR nucleotide sequence can be determined, for example, using the MacDNAsis program (Hitachi Software Engineering America Limited, South San Francisco, CA), Higgins algorithm, with the following exemplary parameters: gap penalty = 5, no. of top diagonals = 5, fixed gap penalty = 5, K-tuple = 2, window size = 5, and floating gap penalty = 10.

D. Generation of Synthetic Env Expression Cassettes

Env coding sequences of the present invention include, but are not limited to, polynucleotide sequences encoding the following HIV-encoded polypeptides: gp160, gp140, and gp120 (see, e.g., U.S. Patent No. 5,792,459 for a description of the HIV-1_{SF2} ("SF2") Env polypeptide). The relationships between these polypeptides is shown schematically in Figure 15 (in the figure: the polypeptides are indicated as lines, the amino and carboxy termini are indicated on the gp160 line; the open circle represents the oligomerization domain; the open square represents a transmembrane spanning domain (TM); and "c" represents the location of a cleavage site, in gp140.mut the "X" indicates that the cleavage site has been mutated such that it no longer functions as a cleavage site). The polypeptide gp160 includes the coding sequences for gp120 and gp41. The polypeptide gp41 is comprised of several domains including an oligomerization domain (OD) and a transmembrane spanning domain (TM). In the native envelope, the oligomerization domain is required for the

non-covalent association of three gp41 polypeptides to form a trimeric structure: through non-covalent interactions with the gp41 trimer (and itself), the gp120 polypeptides are also organized in a trimeric structure.

5 A cleavage site (or cleavage sites) exists approximately between the polypeptide sequences for gp120 and the polypeptide sequences corresponding to gp41. This cleavage site(s) can be mutated to prevent cleavage at the site. The resulting gp140 polypeptide corresponds to

10 a truncated form of gp160 where the transmembrane spanning domain of gp41 has been deleted. This gp140 polypeptide can exist in both monomeric and oligomeric (i.e. trimeric) forms by virtue of the presence of the oligomerization domain in the gp41 moiety. In the

15 situation where the cleavage site has been mutated to prevent cleavage and the transmembrane portion of gp41 has been deleted the resulting polypeptide product is designated "mutated" gp140 (e.g., gp140.mut). As will be apparent to those in the field, the cleavage site can be

20 mutated in a variety of ways. The native amino acid sequence in the SF162 cleavage sites is: APTKAKRRVVQREKR (SEQ ID NO:21), where KAKRR (SEQ ID NO:22) is termed the "second" site and REKR (SEQ ID NO:23) is the "first site". Exemplary mutations include the following

25 constructs: gp140.mut7.modSF162 which encodes the amino acid sequence APTKA**I**SSVVQ**E**K**S** (SEQ ID NO:24) in the cleavage site region; gp140.mut8.modSF162 which encodes the amino acid sequence APTIA**I**SSVVQ**E**K**S** (SEQ ID NO:25) in the cleavage site region and gp140mut.modSF162 which

30 encodes the amino acid sequence APTKAKRRVVQRE**K**S (SEQ ID NO:26). Mutations are denoted in bold. The native amino acid sequence in the US4 cleavage sites is:
APT**Q**A**K**RRVVQREKR (SEQ ID NO:27), where QAKRR (SEQ ID NO:28) is termed the "second" site and REKR (SEQ ID

NO:23) is the "first site". Exemplary mutations include the following construct: gp140.mut.modUS4 which encodes the amino acid sequence APTQAKRRVVQREKS (SEQ ID NO:29) in the cleavage site region. Mutations are denoted in bold.

5

E. Modification of HIV-1 Env (Envelope) Nucleic Acid Coding Sequences

In one embodiment of the present invention, wild-type Env coding sequences were selected from the HIV-1_{SF162} ("SF162") strain (Cheng-Mayer (1989) *PNAS USA* 86:8575-8579). These SF162 sequences were as follows: gp120, SEQ ID NO:30 (Fig. 16); gp140, SEQ ID NO:31 (Fig. 17); and gp160, SEQ ID NO:32 (Fig. 18).

10 In another embodiment of the present invention, wild-type Env coding sequences were selected from the HIV-US4 strain (Mascola, et al. (1994) *J. Infect. Dis.* 169:48-54). These US4 sequences were as follows: gp120, SEQ ID NO:51 (Fig. 38); gp140, SEQ ID NO:52 (Fig. 39); and gp160, SEQ ID NO:53 (Fig. 40).

15 20 These Env coding sequences were manipulated to maximize expression of their gene products.

First, the wild-type coding region was modified in one or more of the following ways. In one embodiment, sequences encoding hypervariable regions of Env, particularly V1 and/or V2 were deleted. In other embodiments, mutations were introduced into sequences encoding the cleavage site in Env to abrogate the enzymatic cleavage of oligomeric gp140 into gp120 monomers. (See, e.g., Earl et al. (1990) *PNAS USA* 87:648-652; Earl et al. (1991) *J. Virol.* 65:31-41). In yet other embodiments, hypervariable region(s) were deleted, N-glycosylation sites were removed and/or cleavage sites mutated.

Second, the HIV-1 codon usage pattern was modified

so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes. The HIV codon usage reflects a high content of the nucleotides A or T in the codon-triplet. The effect of 5 the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The Env coding sequences were modified to be comparable 10 to codon usage found in highly expressed human genes.

Figures 22A-22H present comparisons of the percent A-T content for the cDNAs of stable versus unstable RNAs (comparison window size = 50). Human IFN γ mRNA is known to (i) be unstable, (ii) have a short half-life, and 15 (iii) have a high A-U content. Human GAPDH (glyceraldehyde-3-phosphate dehydrogenase) mRNA is known to (i) be a stable RNA, and (i) have a low A-U content. In Figures 22A-H, the percent A-T content of these two sequences are compared to the percent A-T content of (1) 20 native HIV-1 US4 Env gp160 cDNA, a synthetic US4 Env gp160 cDNA sequence (i.e., having modified codons) of the present invention; and (2) native HIV-1 SF162 Env gp160 cDNA, a synthetic SF162 Env gp160 cDNA sequence (i.e., having modified codons) of the present invention. 25 Figures 22A-H show the percent A-T content over the length of the sequences for IFN γ (Figures 22C and 22G); native gp160 Env US4 and SF162 (Figures 22A and 22E, respectively); GAPDH (Figures 22D and 22H); and the synthetic gp160 Env for US4 and SF162 (Figures 22B and 22F). Experiments performed in support of the present 30 invention showed that the synthetic Env sequences were capable of higher level of protein production (see the Examples) than the native Env sequences. The data in Figures 22A-H suggest that one reason for this increased

production is increased stability of the mRNA corresponding to the synthetic Env coding sequences versus the mRNA corresponding to the native Env coding sequences.

- 5 To create the synthetic coding sequences of the present invention the gene cassettes were designed to comprise the entire coding sequence of interest. Synthetic gene cassettes were constructed by oligonucleotide synthesis and PCR amplification to
- 10 generate gene fragments. Primers were chosen to provide convenient restriction sites for subcloning. The resulting fragments were then ligated to create the entire desired sequence which was then cloned into an appropriate vector. The final synthetic sequences were
- 15 (i) screened by restriction endonuclease digestion and analysis, (ii) subjected to DNA sequencing in order to confirm that the desired sequence had been obtained and (iii) the identity and integrity of the expressed protein confirmed by SDS-PAGE and Western blotting (See,
- 20 Examples. The synthetic coding sequences were assembled at Chiron Corp. or by the Midland Certified Reagent Company (Midland, Texas).

- Exemplary modified coding sequences are presented as synthetic Env expression cassettes in Table 1A and 1B.
- 25 The following expression cassettes (i) have unique, terminal *EcoRI* and *XbaI* cloning sites; (ii) include Kozak sequences to promote optimal translation; (iii) tPA signal sequences (to direct the ENV polypeptide to the cell membrane, see, e.g., Chapman et al., *infra*); (iv)
- 30 open reading frames optimized for expression in mammalian cells; and (v) a translational stop signal codon.

Table 1A: Exemplary Synthetic Env Expression
Cassettes (SF162)

	Expression Cassette	Seq Id	Further Information
5	gp120 SF162	30	wild-type; Figure 16
	gp140 SF162	31	wild-type; Figure 17
	gp160 SF162	32	wild-type; Figure 18
	gp120.modSF162	33	none; Figure 19
10	gp120.modSF162.delV2	34	deleted V2 loop; Figure 20
	gp120.modSF162.delV1/V2	35	deleted V1 and V2; Figure 21
	gp140.modSF162	36	none; Figure 23
	gp140.modSF162.delV2	37	deleted V2 loop; Figure 24
15	gp140.modSF162.delV1/V2	38	deleted V1 and V2; Figure 25
	gp140.mut.modsF162	39	mutated cleavage site; Fig. 26
	gp140.mut.modsF162.delV2	40	deleted V2; mutated cleavage site; Figure 27
	gp140.mut.modsF162.delV1/V2	41	deleted V1 & V2; mutated cleavage site; Figure 28
20	gp140.mut7.modsF162	42	mutated cleavage site; Fig. 29
	gp140.mut7.modsF162.delV2	43	mutated cleavage site; deleted V2; Figure 30
	gp140.mut7.modsF162.delV1/V2	44	mutated cleavage site; deleted V1 and V2; Figure 31
	gp140.mut8.modsF162	45	mutated cleavage site; Fig. 32
25	gp140.mut8.modsF162.delV2	46	mutated cleavage site; deleted V2; Figure 33
	gp140.mut8.modsF162.delV1/V2	47	mutated cleavage site; deleted V1 and V2; Figure 34
	gp160.modsF162	48	none; Figure 35
	gp160.modsF162.delV2	49	deleted V2 loop; Figure 36
	gp160.modsF162.delV1/V2	50	deleted V1 & V2; Figure 37

Table 1B:
Exemplary Synthetic Env Expression Cassettes (US4)

	Expression Cassette	Seq Id	Further Information
5	gp120 US4	51	wild-type; Figure 38
	gp140 US4	52	wild-type; Figure 39
	gp160 US4	53	wild-type; Figure 40
	gp120.modUS4	54	none; Figure 41
	gp120.modUS4.del 128-194	55	deletion in V1 and V2 regions; Figure 42
10	gp140.modUS4	56	none; Figure 43
	gp140.mut.modUS4	57	mutated cleavage site; Figure 44
	gp140TM.modUS4	58	native transmembrane region; Figure 45
	gp140.modUS4.delV1/V2	59	deleted V1 and V2; Figure 46
	gp140.modUS4.delV2	60	deleted V1; Figure 47
15	gp140.mut.modUS4.delV1/V2	61	mutated cleavage site; deleted V1 and V2; Figure 48
	gp140.modUS4.del 128-194	62	deletion in V1 and V2 regions; Figure 49
	gp140.mut.modUS4.del 128-194	63	mutated cleavage site; deletion in V1 and V2 regions; Figure 50
	gp160.modUS4	64	none; Figure 51
	gp160.modUS4.delV1	65	deleted V1; Figure 52
20	gp160.modUS4.delV2	66	deleted V2; Figure 53
	gp160.modUS4.delV1/V2	67	deleted V1 and V2; Figure 54
	gp160.modUS4del 128-194	68	deletion in V1 and V2 regions; Figure 55

Alignments of the sequences presented in the above
25 tables are presented in Figures 66A and 66B.

A common region (Env-common) extends from nucleotide position 1186 to nucleotide position 1329 (SEQ ID NO:69,

Fig. 56) relative to the wild-type US4 sequence and from nucleotide position 1117 to position 1260 (SEQ ID NO:79, Fig. 57) relative to the wild-type SF162 sequence. The synthetic sequences of the present invention
5 corresponding to these regions are presented, as SEQ ID NO:71 (Figure 58) for the synthetic Env US4 common region and as SEQ ID NO:72 (Figure 59) for the synthetic Env SF162 common region.

Percent identity to this sequence can be determined,
10 for example, using the Smith-Waterman search algorithm (Time Logic, Incline Village, NV), with the following exemplary parameters: weight matrix = nuc4x4hb; gap opening penalty = 20, gap extension penalty = 5, reporting threshold = 1; alignment threshold = 20.

15 Various forms of the different embodiments of the present invention (e.g., constructs) may be combined.

F. Cloning Synthetic Env Expression Cassettes of the Present Invention.

20 The synthetic DNA fragments encoding the Env polypeptides were typically cloned into the eucaryotic expression vectors described above for Gag, for example, pCMVKm2/pCMVlink (Figure 4), pCMV6a, pESN2dhfr (Figure 13A), pCMVIII (Figure 13B; alternately designated as the
25 pCMV-PL-E-dhfr/neo vector).

Exemplary designations for pCMVlink vectors containing synthetic expression cassettes of the present invention are as follows: pCMVlink(gp140.modSF162; pCMVlink(gp140.-modSF162.delV2;
30 pCMVlink(gp140.mut.modSF162; pCMVlink(gp140.mut.modSF162.delV2; pCMVKm2(gp140.modUS4; pCMVKm2(gp140.mut.modUS4; and, pCMVKm2(gp140.mut.modUS4.delV1/V2.

G. Generation of Synthetic Tat Expression Cassettes

Tat coding sequences have also been modified according to the teachings of the present specification. The wild type nucleotide sequence encoding tat from 5 variant SF162 is presented in Figure 76 (SEQ ID NO:85). The corresponding wild-type amino acid sequence is presented in Figure 77 (SEQ ID NO:86). Figure 81 (SEQ ID NO:89) shows the nucleotide sequence encoding the amino terminal of the tat protein and the codon encoding cystein-22 is underlined. Other exemplary constructs 10 encoding synthetic tat polypeptides are shown in Figures 78 and 79 (SEQ ID NOs:87 and 88). In one embodiment (SEQ ID NO:88), the cystein residue at position 22 is replaced by a glycine. Caputo et al. (1996) *Gene Therapy* 3:235 15 have shown that this mutation affects the trans activation domain of Tat.

Various forms of the different embodiments of the invention, described herein, may be combined.

20 H. Deposit of Vectors

Selected exemplary constructs shown below and described herein are deposited at Chiron Corporation, Emeryville, CA, 94662-8097, and were sent to the American Type Culture Collection, 10801 University Boulevard, 25 Manassas, VA 20110-2209 on December 27, 1999.

	Plasmid Name	Chiron Deposit #	Date Sent to ATCC
	pCMVgp160.modUS4	5094	27 Dec 99
	pCMVgp160delI.modUS4	5095	27 Dec 99
	pCMVgp160del2.modUS4	5096	27 Dec 99
5	pCMVgp160del-2.modUS4	5097	27 Dec 99
	pCMVgp160del128-194.mod.US4	5098	27 Dec 99
	pCMVgp140mut.modUS4del128-194	5100	27 Dec 99
	pCMVgp140.mut.mod.US	5101	27 Dec 99
	pCMVgp160.modSF162	5125	27 Dec 99
10	pCMVgp160.modSF162.delV2	5126	27 Dec 99
	pCMVgp160.modSF162.delV1V2	5127	27 Dec 99
	pCMVgp140.mut.modSF162delV2	5128	27 Dec 99
	pCMVgp140.mut7.modSF162	5129	27 Dec 99
	pCMVgp140.mut7.modSF162delV2	5130	27 Dec 99
15	pCMVgp140.mut8.modSF162	5131	27 Dec 99
	pCMVgp140.mut8.modSF162delV2	5132	27 Dec 99
	pCMVgp140.mut8.modSF162delV1V2	5133	27 Dec 99
	pCMVKm2.Gagprot.Mod.SF2.GP1	5150	27 Dec 99
	pCMVKm2.Gagprot.Mod.SF2.GP2	5151	27 Dec 99
20			

Example 2Expression Assays for theSynthetic Gag, Env and Tat Coding Sequences25 A. Gag and Gag-Protease Coding Sequences

The HIV-1SF2 wild-type Gag (SEQ ID NO:1) and Gag-protease (SEQ ID NO:2) sequences were cloned into expression vectors having the same features as the vectors into which the synthetic Gag (SEQ ID NO:4) and Gag-protease (SEQ ID NOs:5, 78 or 79)) sequences were cloned.

Expression efficiencies for various vectors carrying the HIV-1SF2 wild-type and synthetic Gag sequences were evaluated as follows. Cells from several mammalian cell lines (293, RD, COS-7, and CHO; all obtained from the 5 American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110-2209) were transfected with 2 µg of DNA in transfection reagent LT1 (PanVera Corporation, 545 Science Dr., Madison, WI). The cells were incubated for 5 hours in reduced serum medium (Opti-MEM, Gibco-BRL, Gaithersburg, MD). The medium was then 10 replaced with normal medium as follows: 293 cells, IMDM, 10% fetal calf serum, 2% glutamine (BioWhittaker, Walkersville, MD); RD and COS-7 cells, D-MEM, 10% fetal calf serum, 2% glutamine (Opti-MEM, Gibco-BRL, 15 Gaithersburg, MD); and CHO cells, Ham's F-12, 10% fetal calf serum, 2% glutamine (Opti-MEM, Gibco-BRL, Gaithersburg, MD). The cells were incubated for either 48 or 60 hours. Supernatants were harvested and filtered through 0.45 µm syringe filters and, optionally, stored 20 at -20°C.

Supernatants were evaluated using the Coulter p24-assay (Coulter Corporation, Hialeah, FL, US), using 96-well plates coated with a murine monoclonal antibody directed against HIV core antigen. The HIV-1 p24 antigen binds to the coated wells. Biotinylated antibodies 25 against HIV recognize the bound p24 antigen. Conjugated strepavidin-horseradish peroxidase reacts with the biotin. Color develops from the reaction of peroxidase with TMB substrate. The reaction is terminated by 30 addition of 4N H₂SO₄. The intensity of the color is directly proportional to the amount of HIV p24 antigen in a sample.

The results of these expression assays are presented in Tables 2A and 2B. Tables 2A and 2B shows data

obtained using the synthetic Gag-protease expression cassette of SEQ ID NO:5. Similar results were obtained using the Gag-protease expression cassettes of SEQ ID NOS:78 and 79.

Table 2: *in vitro* gag and gagprot p24 expression

5 TABLE 2a. Increased *in vitro* expression from modified vs. native gag plasmids in supernatants and lysates from transiently transfected cells

experiment	native (nat) ^a modified (mod) ^b	supernatant (sup) lysate (lys)	cell line	hours post transfection	total ng p24 (fold increase)
1	nat	sup	293	48	3.4
	mod	sup	293	48	1260 (371)
	nat	sup	293	60	3.2
	mod	sup	293	60	2222 (694)
2	nat	sup	293	60	1.8
	mod	sup	293	60	1740 (966)
3	nat	sup	293	60	1.8
	mod	sup	293	60	580 (322)
4	nat	lys	293	60	1.5
	mod	lys	293	60	85 (57)
1	nat	sup	RD	48	5.6
	mod	sup	RD	48	66 (12)
	nat	sup	RD	60	7.8
	mod	sup	RD	60	70.2 (9)
2	nat	lys	RD	60	1.9
	mod	lys	RD	60	7.8 (4)
1	nat	sup	COS-7	48	0.4
	mod	sup	COS-7	48	33.4 (84)
2	nat	sup	COS-7	48	0.4
	mod	sup	COS-7	48	10 (25)
	nat	lys	COS-7	48	3
	mod	lys	COS-7	48	14 (5)

^a pCMVLink.Gag.SF2.PRE

^b pCMVKm2.GagMod.SF2

5 TABLE 2b. *In vitro* expression from modified gag and gagprotease plasmids in supernatants and lysates from transiently transfected cells

plasmid	supernatant (sup) lysate (lys)	cell line	hours post transfection	total ng p24 ^d
Gag ^a	sup	293	60	760
GagProt(GP1) ^b	sup	293	60	380
GagProt(GP2) ^c	sup	293	60	320
Gag	lys	293	60	78
GagProt(GP1)	lys	293	60	1250
GagProt(GP2)	lys	293	60	400
Gag	sup	COS-7	72	40
GagProt(GP1)	sup	COS-7	72	150
GagProt(GP2)	sup	COS-7	72	290
Gag	lys	COS-7	72	60
GagProt(GP1)	lys	COS-7	72	63
GagProt(GP2)	lys	COS-7	72	58

^a pCMVKm2.GagMod.SF2

^b pCMVKm2.GagProtMod.SF2(GP1) gagprotease with codon optimization and inactivation of INS in protease

^c pCMVKm2.GagProtMod.SF2(GP2) gagprotease with only inactivation of INS in protease

^d Shown are representative results from 3 independent experiments for each cell line tested.

The data showed that the synthetic Gag and Gag-protease expression cassettes provided dramatic increases in production of their protein products, relative to the native (HIV-1SF2 wild-type) sequences, when expressed in
5 a variety of cell lines.

B. Env Coding Sequences

The HIV-SF162 ("SF162") wild-type Env (SEQ ID NO:1-3) and HIV-US4 ("US4") wild-type Env (SEQ ID NO:22-24)
10 sequences were cloned into expression vectors having the same features as the vectors into which the synthetic Env sequences were cloned.

Expression efficiencies for various vectors carrying the SF162 and US4 wild-type and synthetic Env sequences
15 were evaluated essentially as described above for Gag except that cell lysates were prepared in 40 μ l lysis buffer (1.0 % NP40, 0.1 M Tris pH 7.5) and frozen at -20°C and capture ELISAs were performed as follows.

For Capture ELISAs, 250 ng of an ammonium sulfate
20 IgG cut of goat polyclonal antibody to gp120SF2/env2-3 was used to coat each well of a 96-well plate (Corning, Corning, NY). Serial dilutions of gp120/SF2 protein (MID 167) were used to set the quantitation curve from which expression of US4 or SF162 gp120 proteins from
25 transfection supernatant and lysates were calculated. Samples were screened undiluted and, optionally, by serial 2-fold dilutions. A human polyclonal antibody to HIV-1 gp120/SF2 was used to detect bound gp120 envelope protein, followed by horse-radish peroxidase (HRP)-
30 labeled goat anti-human IgG conjugates. TMB (Pierce, Rockford, IL) was used as the substrate and the reaction is terminated by addition of 4N H₂SO₄. The reaction was quantified by measuring the optical density (OD) at 450 nm. The intensity of the color is directly

proportional to the amount of HIV gp120 antigen in a sample. Purified SF2 gp120 protein was diluted and used as a standard.

The results of the transient expression assays are
5 presented in Tables 3 and 4. Table 3 depicts transient expression in 293 cells transfected with a pCMVKm2 vector carrying the Env cassette of interest. Table 4 depicts transient expression in RD cells transfected with a pCMVKm2 vector carrying the Env cassette of interest.

5

Table 3

Native (N) Synthetic (S)	Cell Line sup	Total sup (ng)	Sup fold increase (S v. N)	Total lysate (ng)	Cell fold increase (S v. N)	Total (ng)	Total fold increase (S v. N)
N- gp120.US4	RD	87	<1			88	
S- gp120.modus4	RD	690	8	2	5	693	8
N- gp140.US4	RD	526		0		526	
S- gp140.modus4	RD	1305	2	1	2	1306	2
S- gp140mut.modus4	RD	35	N/A	25	N/A	60	N/A
S- gp140TM.modus4	RD	0	N/A	5	N/A	5	N/A
N- gp160.US4	RD	0		8		8	
S- gp160.modus4	RD	0	0	30	4	30	4

Table 4

CHO Cell Lines Expression Level of US4 Envelope Constructs				
	Constructs	CHO Clone #	MTX Level	Expression Level (ng/ml)
5	gp120.modUS4	1	3.2 μ M	250-450
		2	1.6 μ M	350-450
		3	200nM	230-580
		4	200nM	300-500
	gp140.modUS4	1	1 μ M	155-300
		2	1 μ M	100-260
		3	1 μ M	200-430
	gp140.mut. modUS4	1	1 μ M	110-270
		2	1 μ M	100-235
		3	1 μ M	100-220
10	gp140.modUS4 .delV1/V2	1	50nM	313-587**
		2	50nM	237-667**
		3	50nM	492-527**
	gp140.mut. modUS4.delV1 /V2	1	50nM	46-328**
		2	50nM	82-318**
		3	50nM	204-385**

15 *All samples measured at T-75 flask stage unless otherwise indicated

**at 24 well and 6 well plate stages

***in a three liter bioreactor perfusion culture this clone yielded approximately 2-5 μ g/ml.

The data showed that the synthetic Env and expression cassettes provided a significant increase in production of their protein products, relative to the native (HIV-1SF162 or US4 wild-type) sequences, when
5 expressed in a variety of cell lines.

C. CHO Cell line Env expression data

Chinese hamster ovary (CHO) cells were transfected with plasmid DNA encoding the synthetic HIV-1 gp120 or gp140 proteins (e.g., pESN2dhfr or pCMVIII vector backbone) using Mirus TransIT-LT1 polyamine transfection reagent (Pan Vera) according to the manufacturers instructions and incubated for 96 hours. After 96 hours, media was changed to selective media (F12 special with 10 250 µg/ml G418) and cells were split 1:5 and incubated for an additional 48 hours. Media was changed every 5-7 days until colonies started forming at which time the colonies were picked, plated into 96 well plates and screened by gp120 Capture ELISA. Positive clones were 15 expanded in 24 well plates and screened several times for Env protein production by Capture ELISA, as described above. After reaching confluence in 24 well plates, positive clones were expanded to T25 flasks (Corning, Corning, NY). These were screened several times after 20 25 confluence and positive clones were expanded to T75 flasks.

Positive T75 clones were frozen in LN2 and the highest expressing clones amplified with 0-5 µM methotrexate (MTX) at several concentrations and plated in 30 100mm culture dishes. Plates were screened for colony formation and all positive clones were again expanded as described above. Clones were expanded and amplified and screened at each step by gp120 capture ELISA. Positive clones were frozen at each methotrexate level. Highest

producing clones were grown in perfusion bioreactors (3L, 100L) for expansion and adaptation to low serum suspension culture conditions for scale-up to larger bioreactors.

- 5 Tables 5 and 6 show Capture ELISA data from CHO cells transfected with pCMVIII vector carrying a cassette encoding synthetic HIV-US4 and SF162 Env polypeptides (e.g., mutated cleavage sites, modified codon usage and/or deleted hypervariable regions). Thus, stably
10 transfected CHO cell lines which express Env polypeptides (e.g., gp120, gp140-monomeric, and gp140-oligomeric) have been produced.

Table 5

CHO Cell Lines Expression Level of US4 Envelope Constructs				
	Constructs	CHO Clone #	MTX Level	Expression Level* (ng/ml)
5	gp120.modUS4	1	3.2μM	250-450
		2	1.6μM	350-450
		3	200nM	230-580***
		4	200nM	300-500
10	gp140.modUS4	1	1μM	155-300
		2	1μM	100-260
		3	1μM	200-430
15	gp140.mut. modUS4	1	1μM	110-270
		2	1μM	100-235
		3	1μM	100-220
	gp140.modUS4 .delV1/V2	1	50nM	313-587**
		2	50nM	237-667**
		3	50nM	492-527**
	gp140.mut. modUS4.delV1 /V2	1	50nM	46-328**
		2	50nM	82-318**
		3	50nM	204-385**

*All samples measured at T-75 flask stage unless otherwise indicated

**at 24 well and 6 well plate stages

***in a three liter bioreactor perfusion culture this clone yielded approximately 2-5 μg/ml.

Table 6

CHO Cell Lines Expression Level of SF162 Envelope Constructs				
	Constructs	CHO Clone #	MTX Level	Expression Level* (ng/ml)
5	gp120.modSF162	1	0	755-2705
		2	0	928-1538
		3	0	538-1609
	gp140.modSF162	1	20 nM	180-350
	gp140.mut. modSF162	1	20 nM	164-451
		2	20 nM	188-487
		3	20 nM	233-804
	gp120.modSF162 .delV2	1	800nM	528-1560
		2	800nM	487-1878
		3	800nM	589-1212
10	gp140.modSF162 .delV2	1	800nM	300-600
		2	800nM	200-400
		3	800nM	200-500
	gp140.mut. modSF162.delV2	1	800nM	300-700
		2	400nM	1161
		3	800nM	400-600
		4	400nM	1600-2176

15 *All samples measured at T-75 flask stage unless otherwise indicated

20 The results presented above demonstrate the ability of the constructs of the present invention to provide expression of Env polypeptides in CHO cells. Production of polypeptides using CHO cells provides (i) correct glycosylation patterns and protein conformation (as determined by binding to panel of MAbs); (ii) correct binding to CD4 receptor molecules; (iii) absence of non-

mammalian cell contaminants (e.g., insect viruses and/or cells); and (iv) ease of purification.

D. Tat Coding Sequences

5 The HIV-SF162 ("SF162") wild-type Tat (SEQ ID NO:85) sequences were cloned into expression vectors having the same features as the vectors into which the synthetic Tat sequences were cloned (SEQ ID NOS:87, 88 and 89).

10 Expression efficiencies for various vectors carrying the SF162 wild-type and synthetic Tat sequences are evaluated essentially as described above for Gag and Env using capture ELISAs with the appropriate anti-tat antibodies and/or CHO cell assays. Expression of the polypeptides encoded by the synthetic cassettes is
15 improved relative to wild type.

Example 3

Western Blot Analysis of Expression

A. Gag and Gag-Protease Coding Sequences

20 Human 293 cells were transfected as described in Example 2 with pCMV6a-based vectors containing native or synthetic Gag expression cassettes. Cells were cultivated for 60 hours post-transfection. Supernatants were prepared as described. Cell lysates were prepared
25 as follows. The cells were washed once with phosphate-buffered saline, lysed with detergent [1% NP40 (Sigma Chemical Co., St. Louis, MO) in 0.1 M Tris-HCl, pH 7.5], and the lysate transferred into fresh tubes. SDS-polyacrylamide gels (pre-cast 8-16%; Novex, San Diego,
30 CA) were loaded with 20 µl of supernatant or 12.5 µl of cell lysate. A protein standard was also loaded (5 µl, broad size range standard; BioRad Laboratories, Hercules, CA). Electrophoresis was carried out and the proteins were transferred using a BioRad Transfer Chamber (BioRad

Laboratories, Hercules, CA) to Immobilon P membranes (Millipore Corp., Bedford, MA) using the transfer buffer recommended by the manufacturer (Millipore), where the transfer was performed at 100 volts for 90 minutes. The 5 membranes were exposed to HIV-1-positive human patient serum and immunostained using o-phenylenediamine dihydrochloride (OPD; Sigma).

The results of the immunoblotting analysis showed that cells containing the synthetic Gag expression 10 cassette produced the expected p55 protein at higher per-cell concentrations than cells containing the native expression cassette. The Gag p55 protein was seen in both cell lysates and supernatants. The levels of production were significantly higher in cell supernatants 15 for cells transfected with the synthetic Gag expression cassette of the present invention. Experiments performed in support of the present invention suggest that cells containing the synthetic Gag-prot expression cassette produced the expected Gag-prot protein at comparably 20 higher per-cell concentrations than cells containing the native expression cassette.

In addition, supernatants from the transfected 293 cells were fractionated on sucrose gradients. Aliquots of the supernatant were transferred to Polyclear™ ultracentrifuge tubes (Beckman Instruments, Columbia, MD), under-laid with a solution of 20% (wt/wt) sucrose, and subjected to 2 hours centrifugation at 28,000 rpm in a Beckman SW28 rotor. The resulting pellet was suspended in PBS and layered onto a 20-60% (wt/wt) sucrose gradient 25 and subjected to 2 hours centrifugation at 40,000 rpm in a Beckman SW41ti rotor.

The gradient was then fractionated into approximately 10 x 1 ml aliquots (starting at the top, 20%-end, of the gradient). Samples were taken from

fractions 1-9 and were electrophoresed on 8-16% SDS polyacrylamide gels. Fraction number 4 (the peak fraction) corresponds to the expected density of Gag protein VLPs. The supernatants from 293/synthetic Gag 5 cells gave much stronger p55 bands than supernatants from 293/native Gag cells, and, as expected, the highest concentration of p55 in either supernatant was found in fraction 4.

These results demonstrate that the synthetic Gag 10 expression cassette provides superior production of both p55 protein and VLPs, relative to the native Gag coding sequences.

B. Env Coding Sequences

15 Human 293 cells were transfected as described in Example 2 with pCMVKm2-based; pCMVlink-based; p-CMVII-based or pESN2-based vectors containing native or synthetic Env expression cassettes. Cells were cultivated for 48 or 60 hours post-transfection. Cell 20 lysates and supernatants were prepared as described (Example 2). Briefly, the cells were washed once with phosphate-buffered saline, lysed with detergent [1% NP40 (Sigma Chemical Co., St. Louis, MO)] in 0.1 M Tris-HCl, pH 7.5], and the lysate transferred into fresh tubes.

25 SDS-polyacrylamide gels (pre-cast 8-16%; Novex, San Diego, CA) were loaded with 20 μ l of supernatant or 12.5 μ l of cell lysate. A protein molecular weight standard and an HIV SF2 gp120 positive control protein (5 μ l, broad size range standard; BioRad Laboratories, Hercules, 30 CA) were also loaded. Electrophoresis was carried out and the proteins were transferred using a BioRad Transfer Chamber (BioRad Laboratories, Hercules, CA) to Immobilon P membranes (Millipore Corp., Bedford, MA) using the transfer buffer recommended by the manufacturer

(Millipore), where the transfer was performed at 100 volts for 90 minutes. The membranes were then reacted against polyclonal goat anti-gp120SF2/env2-3 anti-sera, followed by incubation with swine anti-goat IgG-peroxidase (POD) (Sigma, St. Louis, MO). Bands indicative of binding were visualized by adding DAB with hydrogen peroxide which deposits a brown precipitate on the membranes.

The results of the immunoblotting analysis showed that cells containing the synthetic Env expression cassette produced the expected Env gp proteins of the predicted molecular weights as determined by mobilities in SDS-polyacrylamide gels at higher per-cell concentrations than cells containing the native expression cassette. The Env proteins were seen in both cell lysates and supernatants. The levels of production were significantly higher in cell supernatants for cells transfected with the synthetic Env expression cassette of the present invention.

20

C. Tat Coding Sequences

Human 293 cells are transfected as described in Example 2 with various vectors containing native or synthetic Tat expression cassettes. Cells are cultivated and isolated proteins analyzed as described above. Immunoblotting analysis shows that cells containing the synthetic Tat expression cassette produced the expected Tat proteins of the predicted molecular weights as determined by mobilities in SDS-polyacrylamide gels at higher per-cell concentrations than cells containing the native expression cassette.

Example 4Purification of Env polypeptidesA. Purification of Oligomeric gp140

Purification of oligomeric gp140 (o-gp140 US4) was conducted essentially as shown in Figure 60. For the experiments described herein, o-gp140 refers to oligomeric gp140 in either native or modified (e.g., optimized expression sequences, deleted, mutated, truncated, etc.) form. Briefly, concentrated (30-50X) supernatants obtained from CHO cell cultures were loaded onto an anion exchange (DEAE) column which removed DNA and other serum proteins. The eluted material was loaded onto a ceramic hydroxyapatite column (CHAP) which bound serum proteins but not HIV Env proteins. The flow-through from the DEAE and CHAP columns was loaded onto a Protein A column as a precautionary step to remove any remaining serum immunoglobulins. The Env proteins in the flow-through were then captured using the lectin *gluvanthus nivalis* (GNA, Vector Labs, Burlingame, CA). GNA has high affinity for mannose rich carbohydrates such as Env. The Env proteins were then eluted with GNA substrate. To remove other highly glycosylated proteins, a cation exchange column (SP) was used to purify gp140/gp120. In a final step, which separates gp120 from o-gp140, a gel filtration column was used to separate oligomers from monomers. Sizing and chromatography analysis of the final product revealed that this strategy lead to the successful isolation of oligomeric gp140.

B. Purification of gp120

Purification of gp120 was conducted essentially as previously described for other Env proteins. Briefly, concentrated supernatants obtained from CHO cell cultures were loaded onto an anion exchange (DEAE) column which

removed DNA and other serum proteins. The eluted material was loaded onto a ceramic hydroxyapatite column (CHAP) which bound serum proteins but not HIV Env proteins. The flow-through from the CHAP column was 5 loaded a cation exchange column (SP) where the flow-through was discarded and the bound fraction eluted with salt. The eluted fraction(s) were loaded onto a Suprose 12/Superdex 200 Tandem column (Pharmacia-Upjohn, Uppsala, Sweden) from which purified gp120 was obtained. Sizing 10 and chromatography analysis of the final product revealed that this strategy successfully purified gp120 proteins.

Example 5

Analysis of Purified Env Polypeptides

15 A. Analysis of o-gp140

It is well documented that HIV Env protein binds to CD4 only in its correct conformation. Accordingly, the ability of o-gp140 US4 polypeptides, produced and purified as described above, to bind CD4 cells was 20 tested. O-gp140 US4 was incubated for 15 minutes with FITC-labeled CD4 at room temperature and loaded onto a Biosil 250 (BioRad) size exclusion column using Waters HPLC. CD4-FITC has the longest retention time (2.67 minutes), followed by CD4-FITC-gp120 (2.167 min). The 25 shortest retention time (1.9 min) was observed for CD4-FITC-o-gp140 US4 indicating that, as expected, o-gp140 US4 binds to CD4 forming a large complex which reduces retention time on the column. Thus, the o-gp140 US4 produced and purified as described above is of the 30 correct size and conformation.

In addition, the US4 o-gp140, purified as described above, was also tested for its ability to bind to a variety of monoclonal antibodies with known epitope specificities for the CD4 binding site, the CD4 inducible

site, the V3 loop and oligomer-specific gp41 epitope. O-gp140 bound strongly to these antibodies, indicating that the purified protein retains its structural integrity.

5 B. Analysis of gp120

As described above, CD4-FITC binds gp120, as demonstrated by the decreased retention time on the HPLC column. Thus, US4 gp120 purified by the above method retains its conformational integrity. In addition, the 10 properties of purified gp120 can be tested by examining its integrity and identity on western blots, as well as, by examining protein concentration, pH, conductivity, endotoxin levels, bioburden and the like. US4 gp120, purified as described above, was also tested for its 15 ability to bind to a variety of monoclonal antibodies with known epitope specificities for the CD4 binding site, the CD4 inducible site, the V3 loop and oligomer-specific gp41 epitope. The pattern of mAb binding to gp120 indicated that the purified protein retained its 20 structural integrity, for example, the purified gp120 did not bind the mAb having the oligomer-specific gp41 epitope (as expected).

Example 6

25 Electron Microscopic Evaluation of VLP Production

The cells for electron microscopy were plated at a density of 50-70% confluence, one day before transfection. The cells were transfected with 10 µg of DNA using transfection reagent LT1 (Panvera) and 30 incubated for 5 hours in serum-reduced medium (see Example 2). The medium was then replaced with normal medium (see Example 2) and the cells were incubated for 14 hours (COS-7) or 40 hours (CHO). After incubation the cells were washed twice with PBS and fixed with 2%

glutaraldehyde. Electron microscopy was performed by Prof. T.S. Benedict Yen, Veterans Affairs, Medical Center, San Francisco, CA).

Electron microscopy was carried out using a
5 transmission electron microscope (Zeiss 10c). The cells were pre-stained with osmium and stained with uranium acetate and lead citrate. The magnification was 100,000X.

Figures 3A and 3B show micrographs of CHO cells
10 transfected with pCMVKM2 carrying the synthetic Gag expression cassette (SEQ ID NO:5) or carrying the Gag-prot expression cassette (SEQ ID NO:79). In the figure, free and budding immature virus-like-particles (VLP) of the expected size (100 nm) are seen for the Gag
15 expression cassette (Figure 3A) and both immature and mature VLPs are seen for the Gag-prot expression cassette (Figure 3B). COS-7 cells transfected with the same vector have the same expression pattern. VLP can also be found intracellularly in CHO and COS-7 cells.

Native and synthetic Gag expression cassettes were compared for their associated levels of VLP production when used to transfect human 293 cells. The comparison was performed by density gradient ultracentrifugation of cell supernatants and Western-blot analysis of the
25 gradient fractions. There was a clear improvement in production of VLPs when using the synthetic Gag construct.

Example 7

30 Expression of Virus-like Particles in the Baculovirus System

A. Expression of Native HIV p55 Gag

To construct the native HIV p55 Gag baculovirus shuttle vector, the prototype SF2 HIV p55 plasmid, pTM1-

Gag (Selby M.J., et al., *J Virol.* 71(10):7827-7831, 1997), was digested with restriction endonucleases *Ncol* and *BamHI* to extract a 1.5 Kb fragment that was subsequently subcloned into pAcC4 (*Bio/Technology* 6:47-55, 1988), a derivative of pAc436. Generation of the recombinant baculovirus was achieved by co-transfected 2 µg of the HIV p55 Gag pAcC4 shuttle vector with 0.5 µg of linearized, *Autographa californica* baculovirus (AcNPV) wild-type viral DNA into *Spodoptera frugiperda* (Sf9) 10 cells (Kitts, P.A., Ayres M.D., and Possee R.D., *Nucleic Acids Res.* 18:5667-5672, 1990). The isolation of recombinant virus expressing HIV p55 Gag was performed according to standard techniques (O'Reilly, D.R., L.K. Miller, and V. A. Luckow, *Baculovirus Expression Vector: A Laboratory Manual*, W.H. Freeman and Company, New York, 1992).

Expression of the HIV p55 Gag was achieved using a 500 ml suspension culture of Sf9 cells grown in serum-free medium (Miaorella, B., D. Inlow, A. Shauger, and D. Harano, *Bio/Technology* 6:1506-1510, 1988) that had been infected with the HIV p55 Gag recombinant baculovirus at a multiplicity of infection (MOI) of 10. Forty-eight hours post-infection, the supernatant was separated by centrifugation and filtered through a 0.2 µm filter. 25 Aliquots of the supernatant were then transferred to Polyclear™ (Beckman Instruments, Palo Alto, CA) ultracentrifuge tubes, underlaid with 20% (wt/wt) sucrose, and subjected to 2 hours centrifugation at 24,000 rpm using a Beckman SW28 rotor.

30 The resulting pellet was suspended in Tris buffer (20 mM Tris HCl, pH 7.5, 250 mM NaCl, and 2.5 mM ethylenediaminetetraacetic acid [EDTA]), layered onto a 20-60% (wt/wt) sucrose gradient, and subjected to 2 hours centrifugation at 40,000 rpm using a Beckman SW41ti

rotor. The gradient was then fractionated starting at the top (20% sucrose) of the gradient into approximately twelve 0.75 ml aliquots. A sample of each fraction was electrophoresed on 8-16% SDS polyacrylamide gels and the 5 resulting bands were visualized after commassie staining (Figure 4). Additional aliquots were subjected to refractive index analysis.

The results shown in Figure 4 indicated that the p55 Gag virus-like particles banded at a sucrose density of 10 range of 1.15 - 1.19 g/ml with the peak at approximately 1.17 g/ml. The peak fractions were pooled and concentrated by a second 20% sucrose pelleting. The resulting pellet was suspended in 1 ml of Tris buffer (described above). The total protein yield as estimated 15 by Bicinchrominic Acid (BCA) (Pierce Chemical, Rockford, IL) was 1.6 mg.

B. Expression of Synthetic HIV p55 Gag

A baculovirus shuttle vector containing the 20 synthetic p55 Gag sequence was constructed as follows. The synthetic HIV p55 expression cassette (Example 1) was digested with restriction enzyme *Sal*I followed by incubation with T4-DNA polymerase. The resulting fragment was isolated (PCR Clean-Up™, Promega, Madison, WI) and then digested with *Bam*HI endonuclease. The 25 shuttle vector pAcC13 (Munemitsu S., et al., *Mol Cell Biol.* 10(11):5977-5982, 1990) was linearized by digestion with *Eco*I, followed by incubation with T4-DNA polymerase, and then isolated (PCR Clean-Up™). The linearized vector 30 was digested with *Bam*HI, treated with alkaline phosphatase, and isolated by size fragmentation in an agarose gel. The isolated 1.5 kb fragment was ligated with the prepared pAcC13 vector. The resulting clone was designated pAcC13-Modif.p55Gag.

The expression conditions for the synthetic HIV p55 VLPs differed from those of the native p55 Gag as follows: a culture volume of 1 liter used instead of 500 ml; *Trichoplusia ni* (Tn5) (Wickham, T.J., and Nermerow, 5 G.R., *BioTechnology Progress*, 9:25-30, 1993) insect cells were used instead of Sf9 insect cells; and, an MOI of 3 was instead of an MOI of 10. Experiments performed in support of the present invention showed that there was no appreciable difference in expression level between the 10 Sf9 and Tn5 insect cells with the native p55 clone. In terms of MOI, experience with the native p55 clone suggested that an MOI of 10 resulted in higher expression (approximately 2-fold) of VLPs than a lower MOI.

The sucrose pelleting and banding methods used for 15 the synthetic p55 VLPs were similar to those employed for the native p55 VLPs (described above), with the following exceptions: pelleted VLPs were suspended in 4 ml of phosphate buffered saline (PBS) instead of 1.0 ml of the Tris buffer; and four, 20-60% sucrose gradients were used 20 instead of a single gradient. Also, due to the high concentration of banded VLPs, further concentration by pelleting was not required. The peak fractions from all 4 gradients were simply dialyzed against PBS. The approximate density of the banded VLPs ranged from 1.23- 25 1.28 g/ml. A total protein yield as estimated by BCA was 46 mg. Results from the sucrose gradient banding of the synthetic p55 are shown in Figure 5.

A comparison of the total amount of purified HIV p55 Gag from several preparations obtained from the two 30 baculovirus expression cassettes has been summarized in Figure 6. The average yield from the native p55 was 3.16 mg/liter of culture (n=5, standard deviation (sd) ± 1.07 , range = 1.8-4.8 mg/L) whereas the average yield from the

synthetic p55 was more than ten-fold higher at 44.5 mg/liter of culture (n=2, sd=±6.4).

In addition to a higher total protein yield, the final product from the synthetic p55-expressed Gag 5 consistently contained lower amounts of contaminating baculovirus proteins than the final product from the native p55-expressed Gag. This difference can be seen in the two commassie-stained gels Figures 4 and 5.

10 C. Expression of Native and Synthetic Gag-Core

Expression of the HIV p55 Gag/HCV Core 173 (SEQ ID NO:8) was achieved using a 2.5 liter suspension culture of Sf9 cells grown in serum-free medium (Miaorella, B., D. Inlow, A. Shauger, and D. Harano. 1988 Bio/Technology 15 6:1506-1510). The cells were infected with an HIV p55 Gag/HCV Core 173 recombinant baculovirus. Forty-eight hours post-infection, the supernatant was separated from the cells by centrifugation and filtered through a 0.2 µm filter. Aliquots of the supernatant were then 20 transferred to a Polyclear™ (Beckman Instruments, Palo Alto, CA) ultracentrifuge tubes containing 30% (wt/wt) sucrose, and subjected to 2 hours of centrifugation at 24,000 rpm in a Beckman SW28 rotor and ultracentrifuge.

The resulting pellet was suspended in Tris buffer 25 (50 mM Tris-HCl, pH 7.5, 500 mM NaCl) and layered onto a 30-60% (wt/wt) sucrose gradient and subjected to 2 hours centrifugation at 40,000 rpm in a Beckman SW41ti rotor and ultracentrifuge. The gradient was then fractionated starting at the top (30%) of the gradient into 30 approximately 11 x 1.0 ml aliquots. A sample of each fraction was electrophoresed on 8-16% SDS polyacrylamide gels and the resulting bands were visualized after commassie staining.

A subset of aliquots were also subjected to Western blot analysis using monoclonal antibody 76C.5EG (Steimer, K.S., et al., *Virology* 150:283-290, 1986) which is specific for HIV p24 (a subunit of HIV p55). The peak 5 fractions from the sucrose gradient were pooled and concentrated by a second 20% sucrose pelleting. The resulting pellet was suspended in 1 ml of buffer Tris buffer and the total protein yield as estimated by BCA (Pierce Chemical, Rockford, IL) was ~ 1.0 mg.

10 The results from the SDS PAGE are shown in Figure 8 and the anti- p24 Western blot results are shown in Figure 9. Taken together, these results indicate that the HIV p55 Gag/HCV Core 173 chimeric VLPs banded at a sucrose density similar to that of the HIV p55 Gag VLPs 15 and the visible protein band that migrated at a molecular weight of ~ 72,000 kd was reactive with the HIV p24-specific monoclonal antibody. An additional immunoreactive band at approximately 55,000 kd also appeared to be reactive with the anti-p24 antibody and 20 may be a degradation product.

Although aliquots from the above preparation were not tested for reactivity with an HCV Core-specific antibody (an anti-CD22 rabbit serum), results from a similar preparation are shown in Figure 10 and indicate 25 that the main HCV Core-specific reactivity migrates at an approximate molecular weight of 72,000 kd which is in accordance with the predicted molecular weight of the chimeric protein.

The expression conditions for the synthetic HIV p55 30 Gag/HCV Core 173 (SEQ ID NO:8) VLPs differed from those of the native p55 Gag and are as follows: a culture volume of 1 liter used instead of 2.5 liters, *Trichoplusia ni* (Tn5) (Wickham, T.J., and Nemerow, G.R. 1993 *BioTechnology Progress*, 9:25-30) insect cells were

used instead of Sf9 insect cells and an MOI of 3 was instead of an MOI of 10. The sucrose pelleting and banding methods used for the synthetic HIV p55 Gag/HCV Core 173 VLPs were similar to those employed for the native HIV p55 Gag/HCV Core 173 VLPs. However, differences included: pelleted VLPs were suspended in 1 ml of phosphate buffered saline (PBS) instead of 1.0 ml of the Tris buffer, and a single 20-60% sucrose gradients was used. A comparison of the total amount of purified HIV p55 Gag/HCV Core 173 from multiple preparations obtained from the two baculovirus expression cassettes showed that there was an increase in expression using the synthetic HIV p55 Gag/HCV Core 173 cassette.

15 D. Alternative method for the enrichment of HIV p55 Gag VLPs

In addition to purification from the media, p55 (Gag protein) expressed in baculovirus (e.g., using a synthetic expression cassette of the present invention) can also be purified as virus-like particles from the infected insect cells. For example, forty-eight hours post infection, the media and cell pellet are separated by centrifugation and the cell pellet is stored at -70°C until future use. At the time of processing, the cell pellet is suspended in 5 volumes of hypotonic lysis buffer (20 mM Tris-HCl, pH 8.2, 1 mM EGTA; 1 mM MgCl₂, and Complete Protease Inhibitor® (Boehringer Mannheim Corp., Indianapolis, IN)). If needed, the cells are then dounced 8-10 times to complete cell lysis.

The lysate is then centrifuged at approximately 1000-1500 x g for 20 minutes. The supernatant is

decanted into UltraClear™ tubes, underlaid with 20% sucrose (w/w) and centrifuged at 24,000 rpm in SW28 buckets for 2 hours. The resulting pellet is suspended in Tris buffer (20 mM Tris HCl, pH 7.5, 250 mM NaCl, and 2.5 mM ethylene-diamine-tetraacetic acid (EDTA) with 0.1% IGEPAL detergent (Sigma Chemical, St. Louis, MO) and 250 units/ml of benzonase (American International Chemical, Inc., Natick, MA) and 10 incubated at 4°C for at least 30 minutes. The suspension is subsequently layered onto a 20-60% sucrose gradient and spun at 40,000 rpm using an SW41ti rotor for 20-24 hours.

After ultracentrifugation, the sucrose gradient is fractionated and aliquots run on SDS PAGE to identify peak fractions. The peak fractions are dialyzed against PBS and measured for protein content. Negatively stained electron micrographs typically show non-enveloped VLPs 20 somewhat smaller in diameter (80-120 nm) than the budded VLPs. HIV Gag VLPs prepared in this manner are also capable of generating Gag-specific CTL responses in mice.

Example 8

25 In Vivo Immunogenicity of Synthetic Gag Expression

Cassettes

A. Immunization

To evaluate the possibly improved immunogenicity of the synthetic Gag expression cassettes, a mouse study was 30 performed. The plasmid DNA, pCMVKM2 carrying the synthetic Gag expression cassette, was diluted to the following final concentrations in a total injection volume of 100 µl: 20 µg, 2 µg, 0.2 µg, and 0.02 µg. To

overcome possible negative dilution effects of the diluted DNA, the total DNA concentration in each sample was brought up to 20 µg using the vector (pCMVKM2) alone. As a control, plasmid DNA of the native Gag expression cassette was handled in the same manner. Twelve groups of four Balb/c mice (Charles River, Boston, MA) were intramuscularly immunized (50 µl per leg, intramuscular injection into the *tibialis anterior*) according to the schedule in Table 7.

10

Table 7

Group	Gag Expression Cassette	Concentration of Gag plasmid DNA (µg)	Immunized at time (weeks):
1	Synthetic	20	0 ¹ , 4
2	Synthetic	2	0, 4
3	Synthetic	0.2	0, 4
4	Synthetic	0.02	0, 4
5	Synthetic	20	0
6	Synthetic	2	0
7	Synthetic	0.2	0
8	Synthetic	0.02	0
9	Native	20	0
10	Native	2	0
11	Native	0.2	0
12	Native	0.02	0

1 = initial immunization at "week 0"

25

Groups 1-4 were bled at week 0 (before immunization), week 4, week 6, week 8, and week 12. Groups 5-12 were bled at week 0 (before immunization) and at week 4.

B. Humoral Immune Response

The humoral immune response was checked with an anti-HIV Gag antibody ELISAs (enzyme-linked immunosorbent assays) of the mice sera 0 and 4 weeks post immunization 5 (groups 5-12) and, in addition, 6 and 8 weeks post immunization, respectively, 2 and 4 weeks post second immunization (groups 1-4).

The antibody titers of the sera were determined by anti-Gag antibody ELISA. Briefly, sera from immunized 10 mice were screened for antibodies directed against the HIV p55 Gag protein. ELISA microtiter plates were coated with 0.2 µg of HIV-1_{SP2} p24-Gag protein per well overnight and washed four times; subsequently, blocking was done with PBS-0.2% Tween (Sigma) for 2 hours. After removal 15 of the blocking solution, 100 µl of diluted mouse serum was added. Sera were tested at 1/25 dilutions and by serial 3-fold dilutions, thereafter. Microtiter plates were washed four times and incubated with a secondary, peroxidase-coupled anti-mouse IgG antibody (Pierce, 20 Rockford, IL). ELISA plates were washed and 100 µl of 3, 3', 5, 5'-tetramethyl benzidine (TMB; Pierce) was added per well. The optical density of each well was measured after 15 minutes. The titers reported are the reciprocal 25 of the dilution of serum that gave a half-maximum optical density (O.D.). The ELISA results are presented in Table 8.

Table 8

Group	Inoculum (μ g)	Expression cassette	Sera - Week 4 ³	Sera - Week 6	Sera - Week 8
5	1	S ¹ - gag	98	455	551
	2	S - gag	59	1408	227
	3	S - gag	29	186	61
	4	S - gag	< 20	< 20	< 20
	5	S - gag	67	n.a. ⁴	n.a.
	6	S - gag	63	n.a.	n.a.
10	7	S - gag	57	n.a.	n.a.
	8	S - gag	< 20	n.a.	n.a.
	9	N ² - gag	43	n.a.	n.a.
	10	N - gag	< 20	n.a.	n.a.
	11	N - gag	< 20	n.a.	n.a.
	12	N - gag	< 20	n.a.	n.a.

1 = synthetic gag expression cassette (SEQ ID NO: 4)

2 = native gag expression cassette (SEQ ID NO: 1)

3 = geometric mean antibody titer

4 = not applicable

20

The results of the mouse immunizations with plasmid-DNAs show that the synthetic expression cassettes provide a clear improvement of immunogenicity relative to the native expression cassettes. Also, the second boost immunization induced a secondary immune response after two weeks (groups 1-3).

25

C. Cellular Immune Response

30

The frequency of specific cytotoxic T-lymphocytes (CTL) was evaluated by a standard chromium release assay of peptide pulsed Balb/c mouse CD4 cells. Gag expressing vaccinia virus infected CD-8 cells were used as a positive control (vvGag). Briefly, spleen cells (Effector cells, E) were obtained from the BALB/c mice immunized as described above (Table 8) were cultured, restimulated, and assayed for CTL activity against Gag

35

peptide-pulsed target cells as described (Doe, B., and Walker, C.M., AIDS 10(7):793-794, 1996). The HIV-1_{SP2} Gag peptide used was p7g SEQ ID NO:10. Cytotoxic activity was measured in a standard ⁵¹Cr release assay. Target (T) 5 cells were cultured with effector (E) cells at various E:T ratios for 4 hours and the average cpm from duplicate wells was used to calculate percent specific ⁵¹Cr release. The results are presented in Table 9.

Cytotoxic T-cell (CTL) activity was measured in 10 splenocytes recovered from the mice immunized with HIV Gag DNA (compare Effector column, Table 9, to immunization schedule, Table 8). Effector cells from the Gag DNA-immunized animals exhibited specific lysis of Gag p7g peptide-pulsed SV-BALB (MHC matched) targets cells 15 indicative of a CTL response. Target cells that were peptide-pulsed and derived from an MHC-unmatched mouse strain (MC57) were not lysed (Table 9; MC/p7g).

Table 9

Table 9. Cytotoxic T-lymphocyte (CTL) responses in mice immunized with HIV-1 gag DNA				
		Percent specific lysis of target cells*		
Immunization	E:T	SVBALB none	SVBALB p7g	RMA p7g
5	20 µg DNA gagmod	100:1	2	49
		30:1	3	30
		10:1	<1	14
	2 µg DNA gagmod	100:1	2	37
		30:1	2	21
		10:1	<1	13
	0.2 µg DNA gagmod	100:1	2	32
		30:1	3	25
		10:1	1	14
10	0.02 µg DNA gagmod	100:1	1	17
		30:1	1	16
		10:1	1	8
	20 µg DNA gag native	100:1	2	49
		30:1	2	24
		10:1	1	12
	2 µg DNA gag native	100:1	<1	18
		30:1	1	14
		10:1	1	7
15	0.2 µg DNA gag native	100:1	3	30
		30:1	3	17
		10:1	2	7
	0.02 µg DNA gag native	100:1	4	2
		30:1	1	2
		10:1	1	2
				<1
				<1
				<1
20				
25				

*representative results of two animals per DNA-dose;

positive CTL responses are indicated by boxed data

The results of the CTL assays show increased potency of synthetic Gag expression cassettes for induction of cytotoxic T-lymphocyte (CTL) responses by DNA immunization.

Example 9In vivo Immunization with Env polypeptidesA. Immunogenicity Study of US4 o-gp140 in Ras-3c Adjuvant System

5 Studies have been conducted using rabbits immunized with US4 o-gp140 purified as described above. Studies are also underway in animals to determine immunogenicity of US4 gp120, SF162 o-gp140 and SF162 gp120.

10 Two rabbits (#1 and #2) were immunized intramuscularly at 0, 4, 12 and 24 weeks with 50 µg of US4 o-gp140 in the Ribi™ adjuvant system (RAS-3c), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphoryllipid A (MPL, Ribi Immunochem, Hamilton, MT).
15 In each experiment described herein, o-gp140 can be native, mutated and/or modified. Antibody responses directed against the US4 o-gp140 protein were measured by ELISA. Results are shown in Table 10.

Table 10

Rabbit/sample	Approximate o-gp140 ELISA titer
pre-immunization	0
#1: post1 (0 week immuniz)	400
5 #1: post2 (4 week immuniz)	15,000
#1: post3 (12 week immuniz)	50,000
#1: post4 (24 week immuiz)	100,000
10 #2: post1 (0 week immuniz)	600
#2: post2 (4 week immuniz)	12,000
#2: post3 (12 week immuniz)	25,000
#2: post4 (24 week immuiz)	55,000

15 The avidities of antibodies directed against the US4 o-gp140 protein were measured in a similar ELISA format employing successive washes with increasing concentrations of ammonium isothiocyanate. Results are shown in Table 11.

20

Table 11

Time of sample	Approx. Antibody avidity (NH ₄ HCN Conc. in M)
pre-immunization	0.02
post1 (0 week immuniz)	1.8
25 post2 (4 week immuniz)	3.5
post3 (12 week immuniz)	5.5
post4 (24 week immuniz)	5.1

30

These results show that US4 o-gp140 is highly immunogenic and able to induce substantial antibody responses after only one or two immunizations.

5 B. Immunogenicity of US4 o-gp140 in MF59-based Adjuvants

Groups of 4 rabbits were immunized intramuscularly at 0, 4, 12 and 24 weeks with various doses of US4 o-gp140 protein in three different MF59-based adjuvants (MF59 is described in International Publication No. WO 90/14837 and typically contains 5% Squalene, 0.5% Tween 80, and 0.5% Span 85). Antibody titers were measured post-third by ELISA using SF2 gp120 to coat the plates. QHC is a quill-based adjuvant (Iscotek, Uppsala, Sweden). Results are shown in Table 12.

15

Table 12

20

25

Antigen dose (μ g)	Adjuvant	Anti-gp120 _{SF2} Ab GMT*
12.5	MF59	7231
25	MF59	8896
50	MF59	12822
12.5	MF59/MPL	24146
25	MF59/MPL	27199
50	MF59/MPL	23059
50	MF59/MPL/QHC	31759

*GMT = geometric mean titer

Thus, adjuvanted o-gp140 generated antigen-specific antibodies. Further, the antibodies were shown to increased in avidity over time.

30

C. Neutralizing Antibodies

Neutralizing antibodies post-third immunization were measured against HIV-1 SF2 in a T-cell line adapted virus

(TCLA) assay and against PBMC-grown HIV-1 variants SF2, SF162 and 119 using the CCR5+ CEMx174 LTR-GFP reporter cell line, 5.25 (provided by N. Landau, Salk Institute, San Diego, CA) as target cells. Results are shown in Table 13.

5

Table 13
Neutralizing antibody responses in rabbits immunized
with o-gp140.modUS4 protein

	Group	Animal	SF2	SF2	SF162	119
			TCLA*	PBMC#	PBMC#	PBMC#
	Experiment 1					
10	o-gp140/ Ras-3c 50 mg	217 218	>640 >640	100% 96	49 37	17 29
15	Experiment 2					
20	o-gp140/ MF59 50 mg	792 793 794 795	45 50 59 128	71 87 87 92	39 26 13 15	26 4 0 0
25	o-gp140/ MF59 + MPL 50 mg	804 805 806 807	173 134 N.D.** 441	91 93 95 100	47 28 49 31	18 4 13 15
30	o-gp140/MF59 + MPL + QHC 50 mg	808 809 810 811	465 496 >640 92	98 100 101 92	46 44 27 24	40 39 4 37

*TCLA neutralizing antibody titers (50% inhibition).

**Not Determined

% Inhibition at 1:10 dilution of sera with any detectable non-specific inhibition in pre-bleeds subtracted.

35

The above studies in rabbits indicate that the US4 o-gp140 protein is highly immunogenic. When administered with adjuvant, this protein was able to induce substantial antibody responses after only one or two immunizations.

5 Moreover, the adjuvanted o-gp140 protein was able to generate antigen-specific antibodies which increased in avidity after successive immunizations, and substantial neutralizing activity against T-cell line adapted HIV-1. Neutralizing activity was also observed against PBMC-grown

10 primary HIV strains, including the difficult to neutralize CCR5 co-receptor (R5)-utilizing isolates, SF162 and 119.

Example 10

In Vivo Immunogenicity of Synthetic Env Expression

15 Cassettes

A. General Immunization Methods

To evaluate the immunogenicity of the synthetic Env expression cassettes, studies using guinea pigs, rabbits, mice, rhesus macaques and baboons were performed. The

20 studies were structured as follows: DNA immunization alone (single or multiple); DNA immunization followed by protein immunization (boost); DNA immunization followed by Sindbis particle immunization; immunization by Sindbis particles alone.

25 B. Humoral Immune Response

The humoral immune response was checked in serum specimens from immunized animals with an anti-HIV Env antibody ELISAs (enzyme-linked immunosorbent assays) at various times post-immunization. The antibody titers of

30 the sera were determined by anti-Env antibody ELISA as described above. Briefly, sera from immunized animals were

screened for antibodies directed against the HIV gp120 or gp140 Env protein. Wells of ELISA microtiter plates were coated

overnight with the selected Env protein and washed four
5 times; subsequently, blocking was done with PBS-0.2% Tween
(Sigma) for 2 hours. After removal of the blocking
solution, 100 μ l of diluted mouse serum was added. Sera
were tested at 1/25 dilutions and by serial 3-fold
dilutions, thereafter. Microtiter plates were washed four
10 times and incubated with a secondary, peroxidase-coupled
anti-mouse IgG antibody (Pierce, Rockford, IL). ELISA
plates were washed and 100 μ l of 3, 3', 5, 5'-tetramethyl
benzidine (TMB; Pierce) was added per well. The optical
density of each well was measured after 15 minutes. Titers
15 are typically reported as the reciprocal of the dilution of
serum that gave a half-maximum optical density (O.D.).

Example 11

DNA-immunization of Baboons Using Synthetic Gag

Expression Cassettes

A. Baboons

Four baboons were immunized 3 times (weeks 0, 4 and 8) bilaterally, intramuscular into the quadriceps using 1mg pCMVKM2.GagMod.SF2 plasmid-DNA (Example 1). The animals
25 were bled two weeks after each immunization and a p24 antibody ELISA was performed with isolated plasma. The ELISA was performed essentially as described in Example 5 except the second antibody-conjugate was an anti-human IgG, g-chain specific, peroxidase conjugate (Sigma Chemical Co.,
30 St. Louis, MD 63178) used at a dilution of 1:500. Fifty μ g/ml yeast extract was added to the dilutions of plasma

samples and antibody conjugate to reduce non-specific background due to

preexisting yeast antibodies in the baboons. The antibody titer results are presented in Table 14.

5

Table 14

	Immunizati on no.	Weeks	Antigen	wpi ^a / Baboon No.	Ab-titer ^b
10	1	0	gagmod DNA	0 w/219	< 10
				0 w/220	< 10
				0 w/221	< 10
				0 w/222	< 10
15	2	6		2 wp 1st/219	< 10
				2 wp 1st/220	< 10
				2 wp 1st/221	< 10
				2 wp 1st/222	15
20	4	14	gagmod DNA	2 wp 4th/219	< 10
				2 wp 4th/220	88
				2 wp 4th/221	< 10
				2 wp 4th/222	56
25	5	30	gagmod DNA	2 wp 5th/219	< 10
				2 wp 5th/220	391
				2 wp 5th/221	237
				2 wp 5th/222	222
30	6	46	gag VLP protein	2 wp 6th/219	753
				2 wp 6th/219	4330
				2 wp 6th/219	5000
				2 wp 6th/219	2881

^a wpi = weeks post immunization

^b geometric mean antibody titer

30

In Table 14, pre-bleed data are given as Immunization No. 0; data for bleeds taken 2 weeks post-first immunization are given as Immunization No. 1; data for bleeds taken 2 weeks post-second immunization are given as Immunization No. 2; and, data for bleeds taken 2 weeks post-third immunization are given as Immunization No. 3.

Further, lymphoproliferative responses to p24 antigen were also observed in baboons 221 and 222 two weeks post-fourth immunization (at week 14), and enhanced substantially post-boosting with VLP (at week 44 and 76).
5 Such proliferation results are indicative of induction of T-helper cell functions.

B. Rhesus Macaques

The improved potency of the codon-modified *gag* expression plasmid observed in mouse and baboon studies was confirmed in rhesus macaques. Four of four macaques had detectable Gag-specific CTL after two or three 1 mg doses of modified *gag* plasmid. In contrast, in a previous study, only one of four macaques given 1 mg doses of plasmid-DNA encoding the wild-type HIV-1_{SF2} Gag showed strong CTL activity that was not apparent until after the seventh immunization. Further evidence of the potency of the modified *gag* plasmid was the observation that CTL from two of the four rhesus macaques reacted with three nonoverlapping Gag peptide pools, suggesting that as many as three different Gag peptides are recognized and indicating that the CTL response is polyclonal. Additional quantification and specificity studies are in progress to further characterize the T cell responses to Gag in the plasmid-immunized rhesus macaques. DNA immunization of macaques with the modified *gag* plasmid did not result in significant antibody responses, with only two of four animals seroconverting at low titers. In contrast, in the same study the majority of macaques in groups immunized with p55Gag protein seroconverted and had strong Gag-specific antibody titers. These data suggest that a prime-boost

strategy (DNA-prime and protein-boost) could be very promising for the induction of a strong CTL and antibody response.

In sum, these results demonstrate that the synthetic
5 Gag plasmid DNA is immunogenic in non-human primates.
When similar experiments were carried out using wild-type
Gag plasmid DNA no such induction of anti-p24 antibodies
was observed after four immunizations.

10

Example 12DNA- and Protein Immunizations of Animals Using Env
Expression Cassettes and PolypeptidesA. Guinea Pigs

Groups comprising six guinea pigs each were
15 immunized intramuscularly at 0, 4, and 12 weeks with
plasmid DNAs encoding the gp120.modUS4, gp140.modUS4,
gp140.modUS4.delV1, gp140.modUS4.delV2,
gp140.modUS4.delV1/V2, or gp160.modUS4 coding sequences
of the US4-derived Env. The animals were subsequently
20 boosted at 18 weeks with a single intramuscular dose of
US4 o-gp140.mut.modUS4 protein in MF59 adjuvant. Anti-
gp120 SF2 antibody titers (geometric mean titers) were
measured at two weeks following the third DNA
immunization and at two weeks after the protein boost.
25 Results are shown in Table 15.

Table 15

Group	GMT post-DNA immuniz.	GMT post-protein boost
gp120.modUS4	2098	9489
gp140.modUS4	190	5340
gp140.modUS4.delV1	341	7808
gp140.modUS4.delV2	386	8165
gp140.modUS4.delV1/V2	664	8270
gp160.modUS4	235	9928

10

These results demonstrate the usefulness of the synthetic constructs to generate immune responses, as well as, the advantage of providing a protein boost to enhance the immune response following DNA immunization.

15

B. Rabbits

Rabbits were immunized intramuscularly and intradermally using a Bioject needless syringe with plasmid DNAs encoding the following synthetic SF162 Env polypeptides: gp120.modSF162, gp120.modSF162.delV2, gp140.modSF162, gp140.modSF162.delV2, gp140.mut.modSF162, gp140.mut.modSF162.delV2, gp160.modSF162, and gp160.modSF162.delV2. Approximately 1 mg of plasmid DNA (pCMVlink) carrying the synthetic Env expression cassette was used to immunize the rabbits. Rabbits were immunized with plasmid DNA at 0, 4, and 12 weeks. At two weeks after the third immunization all of the constructs were shown to have generated significant antibody titers in the test animals. Further, rabbits immunized with constructs containing deletions of the V2 region

generally generated similar antibody titers relative to rabbits immunized with the companion construct still containing the V2 region.

5 The nucleic acid immunizations are followed by protein boosting with o-gp140.modsF162.delV2 (0.1 mg of purified protein) at 24 weeks after the initial immunization. Results are shown in Table 16.

Table 16

Group	GMT 2wks post-2nd DNA immunization	GMT 2wks post-3rd DNA immunization	GMT 2wks post-protein boost
gp120.modsF162	4573	5899	26033
gp120.modsF162.delV2	3811	3122	29606
gp140.modsF162	1478	710	12882
gp140.modsF162.delV2	1572	819	11067
gp140.mut.modsF162	1417	788	8827
15 gp140.mut.modsF162.delV 2	1378	1207	13301
gp160.modsF162	23	81	7050
gp160.modsF162.delV2	85	459	11568

20 All constructs are highly immunogenic and generate substantial antigen binding antibody responses after only 2 immunizations in rabbits.

C. Baboons

25 Groups of four baboons were immunized intramuscularly with 1 mg doses of DNA encoding different forms of synthetic US4 gp140 (see the following table) at 0, 4, 8, 12, 28, and 44 weeks. The animals were also boosted twice with US4 O-gp140 protein (gp140.mut.modUS4) 30 at 44 and 76 weeks using MF59 as adjuvant. Results are shown in Table 17.

Table 17

	Animal	Treatment	2 Wks Post 5th DNA immuniza- tion	2 Wks post 6th DNA (plus o- gp140 prot. immuniz.)	2 Wks post 7th DNA (o-gp140 protein only)
5	CY 215	gp140.modUS4	8.3	446	1813
	CY 216		8.3	433	1236
	CY 217		68	1660	2989
	CY 218		101	2556	1610
10	Geomean:		26.2	951.4	1812.1
	CY 219	gp140.modUS4	8.3	8.3	421
	CY 220		8.3	8.3	3117
	CY 221		8.3	954	871
	CY 222		8.3	71	916
	Geomean:		8.3	46.5	1011.5
15	CY 223	gp140.mut.	41.4	10497	46432
	CY 224		8.3	979	470
	CY 225		135	2935	3870
	CY 226		47	1209	4009
	Geomean:		68.3	2457.4	4289.6
	CY 227	gp140TM.	8.3	56	5001
20	CY 228		8.3	806	1170
	CY 229		8.3	48	3402
	CY 230		8.3	38	6520
	GMT*:		8.3	95.3	3375.3

*GMT = geometric mean titer

The results in Table 17 demonstrate the usefulness
 25 of the synthetic constructs to generate immune responses
 in primates such as baboons. In addition, all animals

showed evidence of antigen-specific (*Env* antigen) lymphoproliferative responses.

D. Rhesus Macaques

5 Two rhesus macaques (designated H445 and J408) were immunized with 1 mg of DNA encoding SF162 gp140 with a deleted V2 region (SF162.gp140.delV2) by intramuscular (IM) and intradermal (ID) routes at 0, 4, 8, and 28 weeks. Approximately 100 µg of the protein encoded by
10 the SF162. gp140mut.delV2 construct was also administered in MF59 by IM delivery at 28 weeks.

ELISA titers are shown in Figure 61. Neutralizing antibody activity is shown Tables 18 and 19.
Neutralizing antibody activity was determined against a
15 variety of primary HIV-1 isolates in a primary lymphocyte or "PBMC-based" assay (see the following tables). Further, the phenotypic co-receptor usage for each of the primary isolates is indicated. As can be seen in the tables neutralizing antibodies were detected against
20 every isolate tested, including the HIV-1 primary isolates (i.e., SF128A, 92US660, 92HT593, 92US657, 92US714, 91US056, and 91US054).

Table 18

	Treatment		Bleed 0	Bleed 1	Bleed 2
Animal	1st Immunization	2nd Immunization	1st Imm'n	2nd Imm'n	2 Wks post 2nd
5	EO 456	25µg 120mod DNA	8.3	45	309
	EO 457		8.3	254	460
	EO 458		8.3	8.3	93
	EO 459		8.3	43	45
	EO 460		8.3	8.3	274
10	EO 461	25µg 120mod DNA	8.3	47	1502
	EO 462		8.3	80	5776
	EO 463		8.3	89	3440
	EO 464		8.3	8.3	3347
	EO 465		8.3	69	1127
15	EO 466	50µg 120mod DNA	8.3	63	102
	EO 467		8.3	112	662
	EO 468		8.3	94	459
	EO 469		8.3	58	48
	EO 470		8.3	95	355
20	EO 471	50µg 120mod DNA	8.3	110	9074
	EO 472		8.3	8.3	4897
	EO 473		8.3	49	4089
	EO 474		8.3	59	5280
	EO 475		8.3	8.3	929
25	EO 476	25µg 120mod DNA	8.3		653
	EO 477		8.3	87	22675
	EO 478		8.3	76	3869
	EO 479		8.3		1004
	EO 480		8.3	71	7080

Table 19

	Treatment		Bleed 0	Bleed 1	Bleed 2
Animal	1st Immunization	2nd Immunization	1st Imm'n	2nd Imm'n	2 Wks post 2nd
EO 481			8.3	8.3	8.3
EO 482			8.3	8.3	8.3
EO 483	Sindbis/Env	(None)	8.3	78	103
EO 484			8.3	8.3	32
EO 485			8.3	76	207
EO 486			8.3	8.3	458
EO 487			8.3	8.3	345
EO 488	Sindbis/Env	Sindbis/Env	8.3	8.3	331
EO 489			8.3	103	111
EO 490			8.3	8.3	5636

Lymphoproliferative activity (LPA) was also determined by antigenic stimulation followed by uptake of ³H-thymidine in these animals and is shown in Table 20. Experiment 1 was performed at 14 weeks post third DNA immunization and Experiment 2 was performed at 2 weeks post fourth DNA immunization using DNA and protein. For gp120ThaiE, gp120SF2 and US4 o-gp140, appropriate background values were used to calculate Stimulation Indices (S.I.; Antigenic stimulation CPM/Background CPM).

10

Table 20

S.I.: Calculated as Ag CPM/Background CPM				
Animal/ exp#	gp120Thai E	gp120 SF2	env2-3SF2	o- gp140US4
J408/#1	2	1	1	5
H445/#1	1	1	1	6
J408/#2	1	1	2	3
H445/#2	0	0	3	2

As can be seen by the results presented in Table 20 lymphoproliferative responses to o-gp140.US4 antigen were also in all four animals at both experimental time points. Such proliferation results are indicative of induction of T-helper cell functions.

The results presented above demonstrate that the synthetic gp140.modSF162.delV2 DNA and protein are immunogenic in non-human primates.

Example 13

In vitro expression of recombinant Sindbis RNA and DNA containing the synthetic Gag or Env expression cassettes

5 A. Synthetic Gag expression cassettes

To evaluate the expression efficiency of the synthetic Gag expression cassette in Alphavirus vectors, the synthetic Gag expression cassette was subcloned into both plasmid DNA-based and recombinant vector particle-based Sindbis virus vectors. Specifically, a cDNA vector construct for *in vitro* transcription of Sindbis virus RNA vector replicons (pRSIN-luc; Dubensky, et al., *J Virol.* 70:508-519, 1996) was modified to contain a *PmeI* site for plasmid linearization and a polylinker for insertion of heterologous genes. A polylinker was generated using two oligonucleotides that contain the sites *XhoI*, *PmlI*, *Apal*, *NarI*, *XbaI*, and *NotI* (XPANXNF, SEQ ID NO:17, and XPANXNR, SEQ ID NO:18).

The plasmid pRSIN-luc (Dubensky et al., *supra*) was digested with *XhoI* and *NotI* to remove the luciferase gene insert, blunt-ended using Klenow and dNTPs, and purified from an agarose gel using GeneCleanII (Biol01, Vista, CA). The oligonucleotides were annealed to each other and ligated into the plasmid. The resulting construct was digested with *NotI* and *SacI* to remove the minimal Sindbis 3'-end sequence and A_{40} tract, and ligated with an approximately 0.4 kbp fragment from PKSSIN1-BV (WO 97/38087). This 0.4 kbp fragment was obtained by digestion of pKSSIN1-BV with *NotI* and *SacI*, and purification after size fractionation from an agarose gel. The fragment contained the complete Sindbis virus 3'-end, an A_{40} tract and a *PmeI* site for linearization. This new vector construct was designated SINBVE.

The synthetic HIV Gag coding sequence was obtained from the parental plasmid by digestion with *Eco*RI, blunting with Klenow and dNTPs, purification with GeneCleanII, digestion with *Sal*I, size fractionation on an agarose gel, and purification from the agarose gel using GeneCleanII. The synthetic Gag coding fragment was ligated into the SINBVE vector that had been digested with *Xho*I and *Pml*I. The resulting vector was purified using GeneCleanII and designated SINBVGag. Vector RNA replicons may be transcribed *in vitro* (Dubensky et al., *supra*) from SINBVGag and used directly for transfection of cells. Alternatively, the replicons may be packaged into recombinant vector particles by co-transfection with defective helper RNAs or using an alphavirus packaging cell line as described, for example, in U.S. Patent Numbers 5,843,723 and 5,789,245, and then administered *in vivo* as described..

The DNA-based Sindbis virus vector pDCMVSIN-beta-gal (Dubensky, et al., *J Virol.* 70:508-519, 1996) was digested with *Sal*I and *Xba*I, to remove the beta-galactosidase gene insert, and purified using GeneCleanII after agarose gel size fractionation. The HIV Gag gene was inserted into the the pDCMVSIN-beta-gal by digestion of SINBVGag with *Sal*I and *Xho*I, purification using GeneCleanII of the Gag-containing fragment after agarose gel size fractionation, and ligation. The resulting construct was designated pDSIN-Gag, and may be used directly for *in vivo* administration or formulated using any of the methods described herein.

BHK and 293 cells were transfected with recombinant Sindbis vector RNA and DNA, respectively. The supernatants and cell lysates were tested with the Coulter p24 capture ELISA (Example 2).

BHK cells were transfected by electroporation with recombinant Sindbis RNA. The expression of p24 (in ng/ml) is presented in Table 21. In the table, SINGag#1 and 2 represent duplicate measurements, and SIN β gal 5 represents a negative control. Supernatants and lysates were collected 24h post transfection.

Table 21

Construct	Supernatant	Lysate
SIN β gal RNA	0	0
SINGag#1 RNA	7 ng	Max (approx. 1 μ g)
SINGag#2 RNA	1 ng	700 ng

293 cells were transfected using LT-1 (Example 2) 15 with recombinant Sindbis DNA. Synthetic pCMVKM2GagMod.SF2 was used as a positive control. Supernatants and lysates were collected 48h post transfection. The expression of p24 (in ng/ml) is presented in Table 22.

20

Table 22

Construct	Supernatant	Lysate
SINGag DNA	3	30
pCMVKM2.GagMod.SF2 DNA	32	42

The results presented in Tables 21 and 22 demonstrate that Gag proteins can be efficiently expressed from both DNA and RNA-based Sindbis vector 30 systems using the synthetic Gag expression cassette (p55Gag.mod).

B. Synthetic Env expression cassettes

To evaluate the expression efficiency of the 35 synthetic Env expression cassette in Alphavirus vectors,

synthetic Env expression cassettes were subcloned into both plasmid DNA-based and recombinant vector particle-based Sindbis virus vectors as described above for Gag.

The synthetic HIV Env coding sequence was obtained 5 from the parental plasmid by digestion with *SalI* and *XbaI*, size fractionation on an agarose gel, and purification from the agarose gel using GeneCleanII. The synthetic Env coding fragment was ligated into the SINBVE vector that had been digested with *XhoI* and *XbaI*. The 10 resulting vector was purified using GeneCleanII and designated SINBVE. Vector RNA replicons may be transcribed *in vitro* (Dubensky et al., *supra*) from SINBVE and used directly for transfection of cells. Alternatively, the replicons may be packaged into 15 recombinant vector particles by co-transfection with defective helper RNAs or using an alphavirus packaging cell line and administered as described above for Gag.

The DNA-based Sindbis virus vector pDCMVSIN-beta-gal (Dubensky, et al., *J Virol.* 70:508-519, 1996) was 20 digested with *SalI* and *XbaI*, to remove the beta-galactosidase gene insert, and purified using GeneCleanII after agarose gel size fractionation. The HIV Env gene was inserted into the the pDCMVSIN-beta-gal by digestion of SINBVE with *XbaI* and *XhoI*, purification using 25 GeneCleanII of the Env-containing fragment after agarose gel size fractionation, and ligation. The resulting construct was designated pDSIN-Env, and may be used directly for *in vivo* administration or formulated using any of the methods described herein.

BHK and 293 cells were transfected with recombinant 30 Sindbis vector RNA and DNA, respectively. The supernatants and cell lysates were tested by capture ELISA.

BHK cells were transfected by electroporation with recombinant Sindbis RNA. The expression of Env (in ng/ml) is presented in Table 23. In the table, the Sindbis RNA containing synthetic Env expression cassettes are indicated and β gal represents a negative control. Supernatants and lysates were collected 24h post transfection.

Table 23

	Construct	Supernatant (Neat) ng/ml	Lysate (1:10 dilution) ng/ml
10	β gal RNA	0	0
15	gp140.modUS4	726	7147
	gp140.modSF162	3529	7772
	gp140.modUS4.delV1/V2	1738	6526
	gp140.modUS4.delV2	960	3023
	gp140.modSF162.delV2	2772	3359

293 cells were transfected using LT-1 mediated transfection (PanVera) with recombinant Sindbis DNA containing synthetic expression cassettes of the present invention and β gal sequences as a negative control. Supernatants and lysates were collected 48h post transfection. The expression of Env (in ng/ml) is presented in Table 24.

Table 24

Construct	Supernatant (Neat) ng/ml	Lysate (1:10 dilution) ng/ml
βgal	0	0
gp140.modSF162.delV2	1977	801
gp140.modSF162	949	746

The results presented in Tables 23 and 24 demonstrated that Env proteins can be efficiently expressed from both DNA and RNA-based Sindbis vector systems using the synthetic Env expression cassettes of the present invention.

Example 14

A. In vivo Immunization with Gag-containing DNA and/or Sindbis particles

CB6F1 mice were immunized intramuscularly at 0 and 4 weeks with plasmid DNA and/or Sindbis vector RNA-containing particles each containing GagMod.SF2 sequences as indicated in Table 25. Animals were challenged with recombinant vaccinia expressing SF2 Gag at 3 weeks post second immunization (at week 7). Spleens were removed from the immunized and challenged animals 5 days later for a standard ⁵¹C release assay for CTL activity. Values shown in Table 25 indicate the results from the spleens of three mice from each group. The boxed values in Table 25 indicate that all groups of mice receiving immunizations with pCMVKm2.GagMod.SF2 DNA and/or SindbisGagMod.SF2 virus particles either alone or in combinations showed antigen-specific CTL activity.

30

Table 25

Cytotoxic T-lymphocyte (CTL) responses in mice immunized with HIV-1 gagmod DNA and Sindbis gagmod virus particles					
		Percent specific lysis of target cells*			
		E:T	SVBALB none	SVBALB p7g	RMA p7g
5	Immunization	100:1	5	20	1
		25:1	5	20	<1
		6:1	4	8	<1
10	SindbisGagMod.SF2 virus particles b at 0, 4 weeks	100:1	10	49	<1
		25:1	7	20	<1
		6:1	5	12	<1
15	pCMVKm2.GagMod.SF2 DNA at 0 wks SindbisGagMod.SF2 virus particles at 4 wks	100:1	9	58	<1
		25:1	7	42	2
		6:1	4	13	<1
20	SindbisGagMod.SF2 virus particles at 4 wks pCMVKm2.GagMod.SF2 DNA at 0 wks	100:1	5	38	<1
		25:1	4	18	<1
		6:1	3	13	1

^a 20 µg
^b 10⁷ particles
* Challenge with recombinant vaccinia virus expressing HIV-1SF2 Gag at 3 weeks post second immunization (week 7). Spleens taken 5 days later. Ex vivo CTL assay performed by standard ⁵¹Cr release assay. Values seen represent results from 3 pooled mouse spleens per group

25

B. In vivo Immunization with Env-containing DNA and/or Sindbis particles

Balb/C mice were immunized intramuscularly at 0 and 4 weeks (as shown in the following table) with plasmid DNA and/or Sindbis-virus RNA-containing particles each containing gp120.modUS4 sequences. Treatment regimes and antibody titers are shown in Table 26. Antibody titers were determined by ELISA using gp120 SF2 protein to coat the plates.

35

Table 26

	Treatment		Bleed 0	Bleed 1 (8 wks)	Bleed 2 (10 wks)	
	Animal	1st Immunization	2nd Immunization	1st Imm'n	2nd Imm'n	2 Wks post 2nd
5	EO 456	25µg 120mod DNA	(None)	8.3	45	309
	EO 457			8.3	254	460
	EO 458			8.3	8.3	93
	EO 459			8.3	43	45
	EO 460			8.3	8.3	274
10	EO 461	25µg 120mod DNA	25µg 120mod DNA	8.3	47	1502
	EO 462			8.3	80	5776
	EO 463			8.3	89	3440
	EO 464			8.3	8.3	3347
	EO 465			8.3	69	1127
15	EO 466	50µg 120mod DNA	(None)	8.3	63	102
	EO 467			8.3	112	662
	EO 468			8.3	94	459
	EO 469			8.3	58	48
	EO 470			8.3	95	355
20	EO 471	50µg 120mod DNA	50µg 120mod DNA	8.3	110	9074
	EO 472			8.3	8.3	4897
	EO 473			8.3	49	4089
	EO 474			8.3	59	5280
	EO 475			8.3	8.3	929
25	EO 476	25µg 120mod DNA	Sindbis/Env	8.3		653
	EO 477			8.3	87	22675
	EO 478			8.3	76	3869
	EO 479			8.3		1004
	EO 480			8.3	71	7080
30	EO 481	Sindbis/Env	(None)	8.3	8.3	8.3
	EO 482			8.3	8.3	8.3
	EO 483			8.3	78	103
	EO 484			8.3	8.3	32
	EO 485			8.3	76	207
35	EO 486	Sindbis/Env	Sindbis/Env	8.3	8.3	458
	EO 487			8.3	8.3	345
	EO 488			8.3	8.3	331
	EO 489			8.3	103	111
	EO 490			8.3	8.3	5636

40 As can be seen from the data presented above, all of the mice generally demonstrated substantial immunological responses by bleed number 2. For Env, the best results were obtained using either (i) 50 µg of gp120.modUS4 DNA for the first immunization followed by a second

immunization using 50 µg of gp120.modUS4 DNA, or (ii) 25 µg of gp120.modUS4 DNA for the first immunization followed by a second immunization using 10⁷ pfus of Sindbis.

5 The results presented above demonstrate that the Env and Gag proteins of the present invention are effective to induce an immune response using Sindbis vector systems which include the synthetic Env (e.g., gp120.modUS4) or Gag expression cassettes.

10

Example 15

Co-Transfection of Env and Gag as Monocistronic and Bicistronic Constructs

DNA constructs encoding (i) wild-type US4 and SF162 Env polypeptides, (ii) synthetic US4 and SF162 Env polypeptides (gp160.modUS4, gp160.modUS4.delV1/V2, gp160.modSF162, and gp120.modSF162.delV2), and (iii) SF2gag polypeptide (i.e., the Gag coding sequences obtained from the SF2 variant or optimized sequences corresponding to the gagSF2 -- gag.modSF2) were prepared. These monocistronic constructs were co-transfected into 293T cells in a transient transfection protocol using the following combinations: gp160.modUS4; gp160.modUS4 and gag.modSF2; gp160.modUS4.delV1/V2; gp160.modUS4.delV1/V2 and gag.modSF2; gp160.modSF162 and gag.modSF2; gp120.modSF162.delV2 and gag.modSF2; and gag.modSF2 alone.

Further several bicistronic constructs were made where the coding sequences for Env and Gag were under the control of a single CMV promoter and, between the two coding sequences, an IRES (internal ribosome entry site (EMCV IRES); Kozak, M., Critical Reviews in Biochemistry and Molecular Biology 27(45):385-402, 1992; Witherell, G.W., et al., Virology 214:660-663, 1995) sequence was

introduced after the Env coding sequence and before the Gag coding sequence. Those constructs were as follows: gp160.modUS4.gag.modSF2, SEQ ID NO:73 (Figure 61); gp160.modUSF162.gag.modSF2, SEQ ID NO:74 (Figure 62);
5 gp160.modUS4.delV1/V2.gag.modSF2, SEQ ID NO:75 (Figure 63); and gp160.modSF162.delV2.gag.modSF2, SEQ ID NO:76 (Figure 64).

Supernatants from cell culture were filtered through 0.45 μ m filters then ultracentrifuged for 2 hours at
10 24,000 rpm (140,000Xg) in an SW28 rotor through a 20% sucrose cushion. The pelleted materials were suspended and layered on a 20-60% sucrose gradient and spun for 2 hours at 40,000 rpm (285,000Xg) in an SW41Ti rotor. Gradients were fractionated into 1.0 ml samples. A total
15 of 9-10 fractions were typically collected from each DNA transfection group.

The fractions were tested for the presence of the Env and Gag proteins (across all fractions). These results demonstrated that the appropriate proteins were
20 expressed in the transfected cells (i.e., if an Env coding sequence was present the corresponding Env protein was detected; if a Gag coding sequence was present the corresponding Gag protein was detected).

Virus like particles (VLPs) were known to be present
25 through a selected range of sucrose densities. Chimeric virus like particles (VLPs) were formed using all the tested combinations of constructs containing both Env and Gag. Significantly more protein was found in the supernatant collected from the cells transfected with
30 "gp160.modUS4.delV1/V2 and gag.modSF2" than in all the other supernatants.

Western blot analysis was also performed on sucrose gradient fractions from each transfection. The results show that bicistronic plasmids gave lower amounts of VLPs

than the amounts obtained using co-transfection with monocistronic plasmids.

In order to verify the production of chimeric VLPs by these cell lines the following electron microscopic analysis was carried out.

293T cells were plated at a density of 60-70% confluence in 100 mm dishes on the day before transfection. The cells were transfected with 10 µg of DNA in transfection reagent LT1 (Panvera Corporation, 545 Science Dr., Madison, WI). The cells were incubated overnight in reduced serum medium (opti-MEM, Gibco-BRL, Gaithersburg, MD). The medium was replaced with 10% fetal calf serum, 2% glutamine in IMDM in the morning of the next day and the cells were incubated for 65 hours. Supernatants and lysates were collected for analysis as described above (see Example 2).

The fixed, transfected 293T cells and purified ENV-GAG VLPs were analyzed by electron microscopy. The cells were fixed as follows. Cell monolayers were washed twice with PBS and fixed with 2% glutaraldehyde. For purified VLPs, gradient peak fractions were collected and concentrated by ultracentrifugation (24,000 rpm) for 2 hours. Electron microscopic analysis was performed by Prof. T.S. Benedict Yen (Veterans Affairs, Medical Center, San Francisco, CA).

Electron microscopy was carried out using a transmission electron microscope (Zeiss 10c). The cells were pre-stained with osmium and stained with uranium acetate and lead citrate. Immunostaining was performed to visualize envelope on the VLP. The magnification was 100,000X.

Figures 65A-65F show micrographs of 293T cells transfected with the following constructs: Figure 65A, gag.modSF2; Figure 65B, gp160.modUS4; Figure 65C,

gp160.modUS4.delV1/V2.gag.modSF2 (bicistronic Env and Gag); Figures 65D and 65E, gp160.modUS4.delV1/V2 and gag.modSF2; and Figure 65F, gp120.modSF162.delV2 and gag.modSF2. In the figures, free and budding immature virus-like-particles (VLPs) of the expected size (approximately 100 nm) decorated with the Env protein were seen. In sum, gp160 polypeptides incorporate into Gag VLPs when constructs were co-transfected into cells. The efficiency of incorporation is 2-3 fold higher when constructs encoding V-deleted Env polypeptides from high synthetic expression cassettes are used.

Although preferred embodiments of the subject invention have been described in some detail, it is understood that obvious variations can be made without departing from the spirit and the scope of the invention as defined by the appended claims.

What Is Claimed Is:

1. An expression cassette, comprising
5 a polynucleotide sequence encoding a polypeptide including an HIV Gag polypeptide, wherein the polynucleotide sequence encoding said Gag polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:20.
10
2. The expression cassette of claim 1, comprising, a polynucleotide sequence encoding a polypeptide including an HIV Gag polypeptide, wherein the polynucleotide sequence encoding said Gag polypeptide
15 comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:9.
3. The expression cassette of claim 1, wherein said polynucleotide sequence encoding a polypeptide including
20 an HIV Gag polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:4.
4. The expression cassette of claim 1, wherein said
25 polynucleotide sequence further includes a polynucleotide sequence encoding an HIV protease polypeptide.
5. The expression cassette of claim 4, wherein the nucleotide sequence encoding said polypeptide comprises a
30 sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:5, SEQ ID NO:78, and SEQ ID NO:79.
6. The expression cassette of claim 1, wherein said

polynucleotide sequence further includes a polynucleotide sequence encoding an HIV *reverse transcriptase polypeptide*.

5 7. The expression cassette of claim 6, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, and SEQ
10 ID NO:84.

8. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HIV *tat polypeptide*.

15 9. The expression cassette of claim 8, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88 and SEQ ID NO:89.

25 10. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HIV *polymerase polypeptide*, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:6.

30 11. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HIV *polymerase polypeptide*, wherein (i) the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90%

sequence identity to the sequence presented as SEQ ID NO:4, and (ii) wherein the sequence is modified by deletions of coding regions corresponding to reverse transcriptase and integrase.

5

12. The expression cassette of claim 11, wherein said polynucleotide sequence preserves T-helper cell and CTL epitopes.

10

13. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HCV core polypeptide, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:7.

15

14. An expression cassette, comprising a polynucleotide sequence encoding a polypeptide including an HIV Env polypeptide, wherein the polynucleotide sequence encoding said Env polypeptide comprises a sequence having at least 90% sequence identity to SEQ ID NO:71 (Figure 58) or SEQ ID NO:72 (Figure 59).

20

15. The expression cassette of claim 14, wherein said Env polypeptide includes sequences flanking a V1 region but has a deletion in the V1 region itself.

25

16. The expression cassette of claim 15, wherein the polynucleotide sequence encoding the polypeptide comprises the sequence presented as SEQ ID NO:65 (Figure 52 gp160.modUS4.delV1).

17. The expression cassette of claim 14, wherein

said Env polypeptide includes sequences flanking a V2 region but has a deletion in the V2 region itself.

18. The expression cassette of claim 17, wherein
5 the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:60 (Figure 47); and SEQ ID NO:66 (Figure 53).

19. The expression cassette of claim 17, wherein
10 the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:34 (Figure 20); SEQ ID NO:37 (Figure 24); SEQ ID NO:40 (Figure 27); SEQ ID NO:43 (Figure 30); SEQ ID NO:46 (Figure 33); SEQ ID NO:49 (Figure 36); and SEQ ID NO:76 (Figure 64).

20. The expression cassette of claim 14, wherein
said Env polypeptide includes sequences flanking a V1/V2 region but has a deletion in the V1/V2 region itself.
20

21. The expression cassette of claim 20, wherein
the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:59 (Figure 46); SEQ ID NO:61 (Figure 48); SEQ ID NO:67 (Figure 54); and SEQ ID NO:75 (Figure 63).

22. The expression cassette of claim 20, wherein
the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:35 (Figure 21); SEQ ID NO:38 (Figure 25); SEQ ID NO:41 (Figure 28); SEQ ID NO:44 (Figure 31); SEQ ID NO:47 (Figure 34) and SEQ ID NO:50 (Figure 37).
30

23. The expression cassette of claim 14, wherein said Env polypeptide has a mutated cleavage site that prevents the cleavage of a gp140 polypeptide into a gp120 polypeptide and a gp41 polypeptide.

5

24. The expression cassette of claim 23, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:57 (Figure 44); SEQ ID NO:61 (Figure 48); and SEQ ID NO:63 (Figure 50).

25. The expression cassette of claim 23, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:39 (Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34).

20

26. The expression cassette of claim 14, wherein said Env polypeptide includes a gp160 Env polypeptide or a polypeptide derived from a gp160 Env polypeptide.

25

27. The expression cassette of claim 26, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:64 (Figure 51); SEQ ID NO:65 (Figure 52); SEQ ID NO:66 (Figure 53); SEQ ID NO:67 (Figure 54); SEQ ID NO:68 (Figure 55); SEQ ID NO:75 (Figure 63); and SEQ ID NO:73 (Figure 61).

30

28. The expression cassette of claim 26, wherein

the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:48 (Figure 35); SEQ ID NO:49 (Figure 36); SEQ ID NO:50 (Figure 37); SEQ ID NO:76 (Figure 64); and SEQ ID NO:74 (Figure 62).

10

29. The expression cassette of claim 14, wherein said Env polypeptide includes a gp140 Env polypeptide or a polypeptide derived from a gp140 Env polypeptide.

30. The expression cassette of claim 29, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:56 (Figure 43); SEQ ID NO:57 (Figure 44); SEQ ID NO:58 (Figure 45); SEQ ID NO:59 (Figure 46); SEQ ID NO:60 (Figure 47); SEQ ID NO:61 (Figure 48); SEQ ID NO:62 (Figure 49); and SEQ ID NO:63 (Figure 50).

20

31. The expression cassette of claim 29, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:36 (Figure 23); SEQ ID NO:37 (Figure 24); SEQ ID NO:38 (Figure 25); SEQ ID NO:39 (Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34).

30

32. The expression cassette of claim 14, wherein said Env polypeptide includes a gp120 Env polypeptide or a polypeptide derived from a gp120 Env polypeptide.

33. The expression cassette of claim 32, wherein
the polynucleotide sequence encoding the polypeptide is
selected from the group consisting of: SEQ ID NO:54
(Figure 41); and SEQ ID NO:55 (Figure 42).

5

34. The expression cassette of claim 32, wherein
the polynucleotide sequence encoding the polypeptide is
selected from the group consisting of: SEQ ID NO:33
(Figure 19); SEQ ID NO:34 (Figure 20); and SEQ ID NO:35
10 (Figure 21).

35. The expression cassette of claim 14, wherein
the polynucleotide sequence encoding the polypeptide is
selected from the group consisting of: SEQ ID NO:55
15 (Figure 42); SEQ ID NO:62 (Figure 49); SEQ ID NO:63
(Figure 50); and SEQ ID NO:68 (Figure 55).

36. A recombinant expression system for use in a
selected host cell, comprising, an expression cassette of
20 any of claims 1-35, and wherein said polynucleotide
sequence is operably linked to control elements
compatible with expression in the selected host cell.

37. The recombinant expression system of claim 36,
25 wherein said control elements are selected from the group
consisting of a transcription promoter, a transcription
enhancer element, a transcription termination signal,
polyadenylation sequences, sequences for optimization of
initiation of translation, and translation termination
30 sequences.

38. The recombinant expression system of claim 36,
wherein said transcription promoter is selected from the

group consisting of CMV, CMV+intron A, SV40, RSV, HIV-Ltr, MMLV-ltr, and metallothionein.

39. A cell comprising an expression cassette of any
5 claims 1-35, and wherein said polynucleotide sequence
is operably linked to control elements compatible with
expression in the selected cell.

40. The cell of claim 39, wherein the cell is a
10 mammalian cell.

41. The cell of claim 40, wherein the cell is
selected from the group consisting of BHK, VERO, HT1080,
293, RD, COS-7, and CHO cells.

15 42. The cell of claim 41, wherein said cell is a
CHO cell.

43. The cell of claim 39, wherein the cell is an
20 insect cell.

44. The cell of claim 43, wherein the cell is
either *Trichoplusia ni* (Tn5) or Sf9 insect cells.

25 45. The cell of claim 39, wherein the cell is a
bacterial cell.

46. The cell of claim 39, wherein the cell is a
yeast cell.

30 47. The cell of claim 39, wherein the cell is a
plant cell.

48. The cell of claim 39, wherein the cell is an antigen presenting cell.

49. The cell of claim 48, wherein the lymphoid cell
5 is selected from the group consisting of macrophage,
monocytes, dendritic cells, B-cells, T-cells, stem cells,
and progenitor cells thereof.

50. The cell of claim 39, wherein the cell is a
10 primary cell.

51. The cell of claim 39, wherein the cell is an immortalized cell.

15 52. The cell of claim 39, wherein the cell is a tumor-derived cell.

53. A method for producing a polypeptide including
HIV Gag polypeptide sequences, said method comprising,
20 incubating the cells of claim 39, under conditions
for producing said polypeptide.

54. A method for producing virus-like particles (VLPs), comprising,
25 incubating the cells of claim 39, under conditions
for producing said VLPs.

55. A method for producing a composition of virus-like particles (VLPs), comprising,
30 (a) incubating the cells of claim 39, under conditions for producing said VLPs; and
(b) substantially purifying said VLPs to produce a composition of VLPs.

56. A cell line useful for packaging lentivirus vectors, comprising

suitable host cells that have been transfected with an expression vector containing an expression cassette of
5 any of claims 1-35, and wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the host cell.

10 57. The cell line of claim 56, wherein suitable host cells have been transfected with an expression vector containing the expression cassette of any of claims 1-13.

15 58. The cell line of claim 56, wherein suitable host cells have been transfected with an expression vector containing the expression cassette of claim 1-3.

20 59. The cell line of claim 56, wherein suitable host cells have been transfected with an expression vector containing the expression cassette of claim 14-35.

60. A gene delivery vector for use in a Mammalian subject, comprising

25 a suitable gene delivery vector for use in said subject, wherein the vector comprises an expression cassette of any of claims 1-35, and wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the subject.

30 61. A method of DNA immunization of a subject, comprising,

introducing a gene delivery vector of claim 60 into said subject under conditions that are compatible with expression of said expression cassette in said subject.

62. The method of claim 61, wherein said gene delivery vector is a nonviral vector.

5 63. The method of claim 61, wherein said vector is delivered using a particulate carrier.

10 64. The method of claim 63, wherein said vector is coated on a gold or tungsten particle and said coated particle is delivered to said subject using a gene gun.

65. The method of claim 63, wherein said vector is encapsulated in a liposome preparation.

15 66. The method of claim 61, wherein said vector is a viral vector.

67. The method of claim 66, wherein said viral vector is a retroviral vector.

20 68. The method of claim 67, wherein said viral vector is a lentiviral vector.

25 69. The method of claim 61, wherein said subject is a mammal.

70. The method of claim 69, wherein said mammal is a human.

30 71. A method of generating an immune response in a subject, comprising
transfected cells of said subject a gene delivery vector of claim 60, under conditions that permit the expression of said polynucleotide and production of said

polypeptide, thereby eliciting an immunological response to said polypeptide.

5 72. The method of claim 71, wherein said vector is a nonviral vector.

73. The method of claim 72, wherein said vector is delivered using a particulate carrier.

10 74. The method of claim 73, wherein said vector is coated on a gold or tungsten particle and said coated particle is delivered to said vertebrate cell using a gene gun.

15 75. The method of claim 73, wherein said vector is encapsulated in a liposome preparation.

76. The method of claim 71, wherein said vector is a viral vector.

20 77. The method of claim 76, wherein said viral vector is a retroviral vector.

25 78. The method of claim 77, wherein said viral vector is a lentiviral vector.

79. The method of claim 71, wherein said subject is a mammal.

30 80. The method of claim 79, wherein said mammal is a human.

81. The method of claim 71, wherein said transfecting is done *ex vivo* and said transfected cells

are reintroduced into said subject.

82. The method of claim 71, wherein said transfecting is done *in vivo* in said subject.

5

83. The method of claim 71, where said immune response is a humoral immune response.

84. The method of claim 71, where said immune
10 response is a cellular immune response.

85. A gene delivery vector comprising an alphavirus vector construct, wherein said alphavirus construct comprises an expression cassette according to any one of
15 claims 1 through 35.

86. The gene delivery vector of claim 85, wherein the alphavirus vector construct is a cDNA vector construct.

20

87. The gene delivery vector of claim 85, wherein the alphavirus comprises a recombinant alphavirus particle preparation.

25

88. The gene delivery vector of claim 85, wherein the vector comprises a eukaryotic layered vector initiation system.

30

89. A method of stimulating an immune response in a subject comprising administering the gene delivery vector of any one of claims 85 through 88 in an amount effective to stimulate an immune response in said subject.

90. The method of claim 89, wherein the gene

delivery vector is administered intramuscularly, intramucosally, intranasally, subcutaneously, intradermally, transdermally, intravaginally, intrarectally, orally or intravenously.

orig.gagSF2

ATGGGTGCGAGAGCGTCGGTATTAAGCGGGGAGAATTAGATAAATGGAAAAATTGGTTAAGGCCAGGGGAAAG

Inact. 1
AAAAAAATATAAGTTAAAACATATA GTATGGCAAGCAGGGAGCTAGAACGATTGCAGTCATCCTGGCCTGTTAGAA
G G C C G C C

Inact. 2
ACATCAGAAGGCTGCAGACAAATATTGGACAGCTACAGCCATCCCTTCAGACAGGATCAGAGAACCTAGATCATTAA
G G C C

Inact. 3
TATAATACAGTAGCAACCCCTCTATTGTGTACATCAAAGGATAGATGTAAGAACACCAAGGAAGCTTAGAGAAAGATA
G C C C G

Inact. 4
GAGGAAGAGCAAAACANAAAGTAAGAAAAGGCACAGCAAGCAGCAGCTGCAGCTGGCACAGGAAACAGCAGCCAGGTC
GTCC G C G

AGCCAAAATTACCCATAGTGCAGAACCTACAGGGCAATGGTACATCAGGCCATATCACCTAGAACTTAAATGCA

TGGGTAAAAGTAGTAGAAGAAAAGGCTTCAGCCCAGAAGTAATACCCATGTTTCAGCATTATCAGAAGGAGCCACC

Inact. 5
CCACAGATTTAACACCATGCTAAACACAGTGCCCCACATCAAGCAGCCATGCAAATGTTAAAAGAGACTATCAAT
G CC G G T G C

GAGGAAGCTGCAGAATGGGATAGAGTGCATCCAGTGCAGGGCTATTGCACCAGGCCAATGAGAGAACCAAGG

GGAAAGTGCACATAGCAGGAACACTAGTACCCCTCAGGAACAAATAGGATGGATGACAATAATCCACCTATCCCAGTA

Inact. 6
GGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAAAATAGTAAGAATGTATAGCCCTACCAGCATTCTGGAC
G C G G G

Inact. 7
ATAAGACAAGGACCAAGGAACCCTTAGAGATTATGTAGACCGGTTCTATAAAACTCTAAGAGGCGAACAGCTTCA
G G

Inact. 8
CAGGATGTTAAAATTGGATGACAGAACCTTGGTCCAAATGCAAACCCAGATTGTAAGACATTGAGAGAACCAAGG
C C G G T

TTGGGACCCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGACCCGCCATAAGCAAGAGTT
C C G

TTGGCTGAAGCCATGAGCCAAGTAACAAATCCAGCTAACATAATGATGCAGAGAGGCAATTAGGAACCAAGAAAG
ACTGTTAAGTGTTCATTGGCAAAGAAGGGCACATAGCCAAAATTGCAGGGCCCTAGGAAAAGGGCTGTTGG

AGATGTGGAAGGGAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTAGGAAGATCTGGCCTTCC

TACAAGGGAGGCCAGGGATTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTCAGGTTGGG

GAGGAGAAAACAACCTCCCTCAGAAGCAGGAGGCCATAGACAAGGAACGTATCCTTAACCTCCCTCAGATCACTC

TTTGGCAACGACCCCTCGTCACAATAA

FIG. 1

native HIV-1SF2 gag-protease

From here codon optimization + inactivation (GP1) and (GP2)

ATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATAAAATGGGAAAAAAATT CGGTTAAGGCCAGGGGGAAAG

Inact. 1
A A A A A T A T A G T T A A A C A T A T G T A T G G G C A A G C A G G G A G C T A G A A C G A T T C G C A G T C A A T C C T G G C T G T T A G A A
G G C C C G C G

Inact. 2 TATAATA **Inact. 3** CAGTAGCAACCTCTATTGTGTACATCAAAGGATAGATGTAAAA GACACCCAAGGAAGCTTCTAGAGAAGATA

GAGGAAGAGCAAAACAAAGTAAGGGCACAGCAGCTGCAGCTGGCACAGGAAACAGCAGCCAGGTC
Inact. 4
GTCC G C G

AGCCAAAATTACCTATAGTGCAGAACCTACAGGGGCAATGGTACATCAGGCCATACCTACCTAGAACTTTAATGCA

TGGGTAAAAAGTAGTGAAGAAAAGGCTTCAGCCCCAGAAGTAATACCCATGTTTCAGCATTATCAGAAGGAGGCCACC

CCACAGATTTAACACCATGCTAACACAGTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGAGACTATCAAT
G CC G G T G

GAGGAAGCTGCAGAATGGGATAAGAGTGCATCCAGTCAGGGCCTATTGCACCAAGGCCAAATGAGAGAACCAAGG

Insert 6

ATAAGACAAGGACCAAGGAACCCTTAGAGATTATGTAGACCGGTTCTATAAAACTCTAAGAGCGAACAAAGCTTC

CAGGATGTAAAAAAATTGGATGACAGAAACCTTGGTCCAAATGCAACCCAGATTGTAAGACTTTTAAAAGCA
Inact. 7
S C G G T

Inact.⁷
TTGGGACAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAAGGGAGTGGGGGGACCCGGCCATAAGCAAGAGTT
S S S G

Inact. 8	Inact. 9	Inact. 10
TTGGGTGAAGCCATGAGCCAGTACAAATCCAGCTAACATAATGATGCAGAGAGGCAATTAGGAACCAAAGAAG C G G G G C G C	ACTGTTAAGTGTTCATTGTGGCAAAGAAGGGCACATAGCCAAAAATTGCAAGGGCCCCTAGGAAAGGGCTGTTGG G G G G G C G C	

ACATCTCCGAGGGAGGGACACCAATTGAAAGATTGACTGAGAGACAGGCTAATTTTTAGGGAAGATCTGGCCTT

TACAAGGGAAGGCCAGGGATTTCAGACGACAGAGCCAAAGCCCCACCAGAAGAGAGCTTCAGGTTTGG

GAGGAGAAAACAACCTCCCTCTCAGAACGCAGGAGCCGATAGACAAGGA
From here codon optimization + inactivation (GP1)

Inact. II of Only Inactivation (G77)
TTTGGCAACGACCCCTCGTCACATAAGGATAGGGGGCAACTAAAGGAAGCTCTATTGATACAGGAGCAGATGATA
G C C G GC C

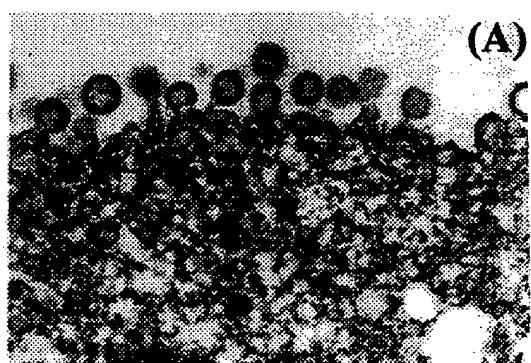
Inact. 12 CAGTATTAGAAGAAATGAATTGCCAGGAAATGGAAACCAAAATGATAAGGGGAAATTGGAGGTTTATCAAAGTA
G C G C C G G

Inact.13 Inact.14
ACATAATTGGAAGAAATCTGTTGACTCAGATTGGTTGACTTTAAATTCCCCATTAGTCCTATTGAAACTGTACCA
G C S C G C C C G G C

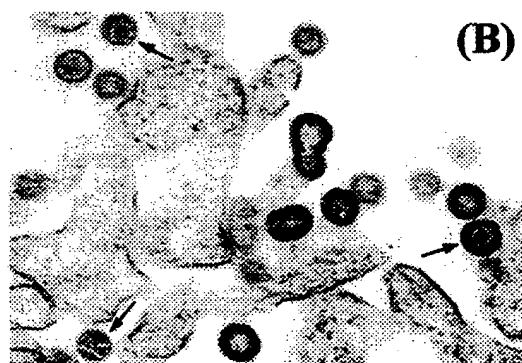
Inact. 14
TAAATTAAAGCCAGGAATGGATGGCCCCAAAGTTAAGCAATGCCATTGTAAGCTGCGCGC
G G G G C C G G

FIG. 2

FIG. 2



(A)



(B)

FIG. 3A

FIG. 3B

4 / 131

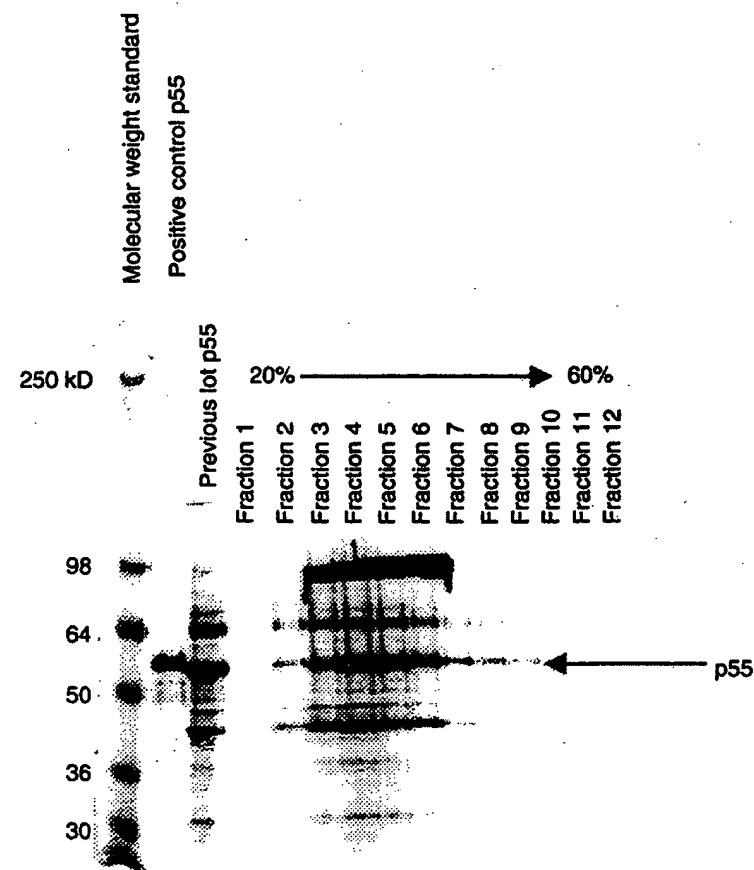


FIG. 4

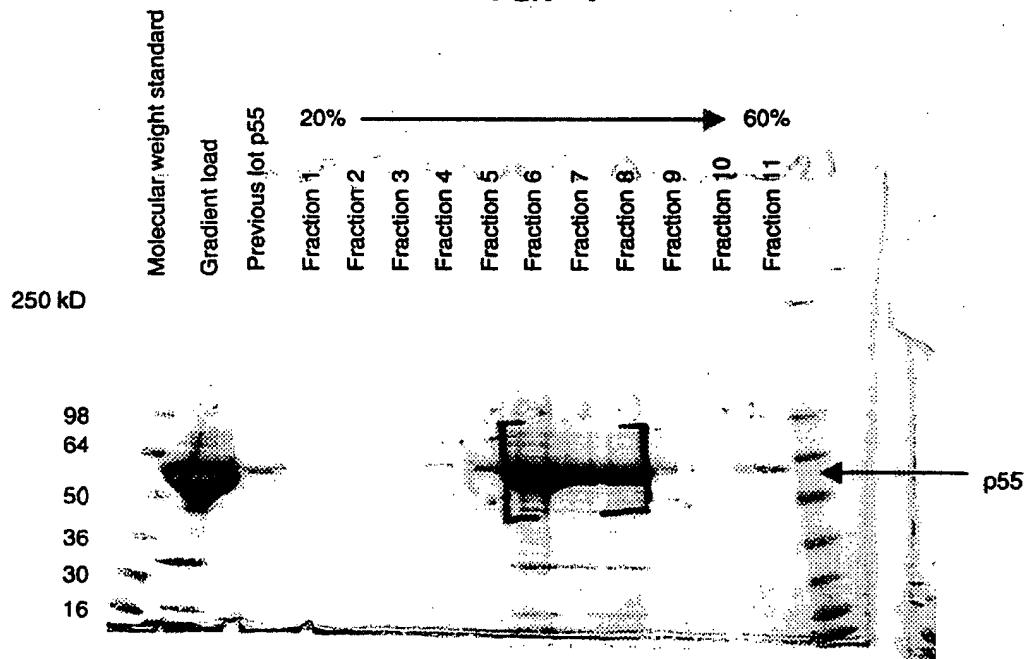


FIG. 5

5 / 131

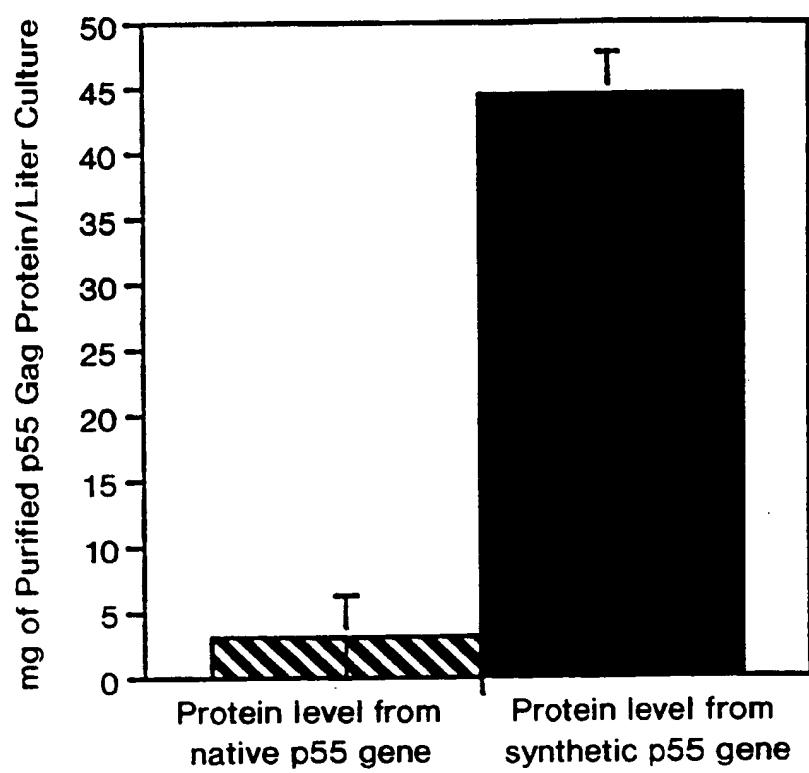


FIG. 6

		10	20	30	40	50
GagPol.ModSF	1 ATGGCGCCC	GCGCCAGCGT	GCTGAGGGC	GGCGAGCTGG	ACAAGTGGGA	50
GagProt.Mods	1 ATGGGGCCC	GGGCCAGCGT	GCTGAGGGC	GGCGAGCTGG	ACAAGTGGGA	50
Gag.ModSF2	1 ATGGGCGCC	GCGCCAGCGT	GCTGAGGGC	GGCGAGCTGG	ACAAGTGGGA	50
GagPol.ModSF	51 GAAGATCCGC	CTGCGCCCCG	GGGCAAGAA	GAAGTACAG	CTGAAGCAC	100
GagProt.Mods	51 GAAGATCCGC	CTGCGCCCCG	GGGCAAGAA	GAAGTACAG	CTGAAGCAC	100
Gag.ModSF2	51 GAAGATCCGC	CTGCGCCCCG	GGGCAAGAA	GAAGTACAG	CTGAAGCAC	100
GagPol.ModSF	101 TCGTGTGGC	CAGCCCGAG	CTGGAGCGCT	TGCGCGTGA	CCCCGGCTG	150
GagProt.Mods	101 TCGTGTGGC	CAGCCCGAG	CTGGAGCGCT	TGCGCGTGA	CCCCGGCTG	150
Gag.ModSF2	101 TCGTGTGGC	CAGCCCGAG	CTGGAGCGCT	TGCGCGTGA	CCCCGGCTG	150
GagPol.ModSF	151 CTGGAGACCA	GCGAGGGCTG	CGCAGATC	CTGGCCAGTC	TGCAGCCCCAG	200
GagProt.Mods	151 CTGGAGACCA	GCGAGGGCTG	CGCAGATC	CTGGCCAGTC	TGCAGCCCCAG	200
Gag.ModSF2	151 CTGGAGACCA	GCGAGGGCTG	CGCAGATC	CTGGCCAGTC	TGCAGCCCCAG	200
GagPol.ModSF	201 CCTGAGACC	GGCAGCGAGG	AGCTGCGCAG	CCTGTACAC	ACCGTGGCCA	250
GagProt.Mods	201 CCTGAGACC	GGCAGCGAGG	AGCTGCGCAG	CCTGTACAC	ACCGTGGCCA	250
Gag.ModSF2	201 CCTGAGACC	GGCAGCGAGG	AGCTGCGCAG	CCTGTACAC	ACCGTGGCCA	250
GagPol.ModSF	251 CCCTGTACTG	C GTGACCCAG	CGCATCGACG	TCAAGGACAC	CAAGGAGGCC	300
GagProt.Mods	251 CCCTGTACTG	C GTGACCCAG	CGCATCGACG	TCAAGGACAC	CAAGGAGGCC	300
Gag.ModSF2	251 CCCTGTACTG	C GTGACCCAG	CGCATCGACG	TCAAGGACAC	CAAGGAGGCC	300
GagPol.ModSF	301 CTGGAGAGA	T CGAGGAGGA	GCAGAACAAAG	TCCAAGAAGA	AGGCCAGCA	350
GagProt.Mods	301 CTGGAGAGA	T CGAGGAGGA	GCAGAACAAAG	TCCAAGAAGA	AGGCCAGCA	350
Gag.ModSF2	301 CTGGAGAGA	T CGAGGAGGA	GCAGAACAAAG	TCCAAGAAGA	AGGCCAGCA	350
GagPol.ModSF	351 GGGCGCCGCC	GCGCCGGCA	CGGGCAACAG	CAGCCAGGTG	AGCCAGAACT	400
GagProt.Mods	351 GGGCGCCGCC	GCGCCGGCA	CGGGCAACAG	CAGCCAGGTG	AGCCAGAACT	400
Gag.ModSF2	351 GGGCGCCGCC	GCGCCGGCA	CGGGCAACAG	CAGCCAGGTG	AGCCAGAACT	400
GagPol.ModSF	401 ACCCCATCGT	GCAGAACCTG	CAGGGCCAGA	TGGTGCACCA	GGCCATCAGC	450
GagProt.Mods	401 ACCCCATCGT	GCAGAACCTG	CAGGGCCAGA	TGGTGCACCA	GGCCATCAGC	450
Gag.ModSF2	401 ACCCCATCGT	GCAGAACCTG	CAGGGCCAGA	TGGTGCACCA	GGCCATCAGC	450

FIG. 7A

7 / 131

FIG. 7B

GagPol. ModSF	451	CCCCGCACCC	460	470	480	490	500
GagProt. ModS	451	CCCCGCACCC	TGAAACGCCCTG	GGTGAAGGTG	GTGGAGGAGA	AGGCCTTCAG	500
Gag. ModSF2	451	CCCCGCACCC	TGAAACGCCCTG	GGTGAAGGTG	GTGGAGGAGA	AGGCCTTCAG	500
GagPol. ModSF	501	CCCCGAGGTG	ATCCCCATGT	TCAGGCCCT	GAGCGAGGGC	GCCACCCCCC	550
GagProt. ModS	501	CCCCGAGGTG	ATCCCCATGT	TCAGGCCCT	GAGCGAGGGC	GCCACCCCCC	550
Gag. ModSF2	501	CCCCGAGGTG	ATCCCCATGT	TCAGGCCCT	GAGCGAGGGC	GCCACCCCCC	550
GagPol. ModSF	551	AGGACCTGAA	CACGATGTG	AACACCGTGG	GGGGCCACCA	GGGCCCATG	600
GagProt. ModS	551	AGGACCTGAA	CACGATGTG	AACACCGTGG	GGGGCCACCA	GGGCCCATG	600
Gag. ModSF2	551	AGGACCTGAA	CACGATGTG	AACACCGTGG	GGGGCCACCA	GGGCCCATG	600
GagPol. ModSF	601	CAGATGCTGA	AGGAGACCAT	CAACGAGGAG	GGCCCGAGT	GGGACCGCGT	650
GagProt. ModS	601	CAGATGCTGA	AGGAGACCAT	CAACGAGGAG	GGCCCGAGT	GGGACCGCGT	650
Gag. ModSF2	601	CAGATGCTGA	AGGAGACCAT	CAACGAGGAG	GGCCCGAGT	GGGACCGCGT	650
GagPol. ModSF	651	GCACCCCCGTG	CACGCCGGCC	CCATCGCCCC	CGCCCAAGATG	CGCGAGCCCC	700
GagProt. ModS	651	GCACCCCCGTG	CACGCCGGCC	CCATCGCCCC	CGCCCAAGATG	CGCGAGCCCC	700
Gag. ModSF2	651	GCACCCCCGTG	CACGCCGGCC	CCATCGCCCC	CGCCCAAGATG	CGCGAGCCCC	700
GagPol. ModSF	701	GGGCAGCGCA	CATCGCCGGC	ACCACCAAGCA	CCCTGCAGGA	GCAGATCGGC	750
GagProt. ModS	701	GGGCAGCGCA	CATCGCCGGC	ACCACCAAGCA	CCCTGCAGGA	GCAGATCGGC	750
Gag. ModSF2	701	GGGCAGCGCA	CATCGCCGGC	ACCACCAAGCA	CCCTGCAGGA	GCAGATCGGC	750
GagPol. ModSF	751	TGGATGACCA	ACAACCCCC	CATCCCCGTG	GGCGAGATCT	ACAAGCGGTG	800
GagProt. ModS	751	TGGATGACCA	ACAACCCCC	CATCCCCGTG	GGCGAGATCT	ACAAGCGGTG	800
Gag. ModSF2	751	TGGATGACCA	ACAACCCCC	CATCCCCGTG	GGCGAGATCT	ACAAGCGGTG	800
GagPol. ModSF	801	GATCATCCTG	GGCCTGAAACA	AGATCGTGC	GATGTACAGC	CCCACCAAGCA	850
GagProt. ModS	801	GATCATCCTG	GGCCTGAAACA	AGATCGTGC	GATGTACAGC	CCCACCAAGCA	850
Gag. ModSF2	801	GATCATCCTG	GGCCTGAAACA	AGATCGTGC	GATGTACAGC	CCCACCAAGCA	850
GagPol. ModSF	851	TCCCTGGACAT	CCGCCAGGGC	CCCAAGGGC	CCTTCCGGCA	CTACGTGGAC	900
GagProt. ModS	851	TCCCTGGACAT	CCGCCAGGGC	CCCAAGGGC	CCTTCCGGCA	CTACGTGGAC	900
Gag. ModSF2	851	TCCCTGGACAT	CCGCCAGGGC	CCCAAGGGC	CCTTCCGGCA	CTACGTGGAC	900

FIG. 7C

GagPol.ModSF	901	CGCTTCTACA	AGACCCCTGCG	CGCTGAGCAG	GCCAGCCAGG	ACGTGAAGAA	950
GagProt.ModS	901	CGCTTCTACA	AGACCCCTGCG	CGCTGAGCAG	GCCAGCCAGG	ACGTGAAGAA	950
Gag.ModSF2	901	CGCTTCTACA	AGACCCCTGCG	CGCTGAGCAG	GCCAGCCAGG	ACGTGAAGAA	950
GagPol.ModSF	951	CTGGATGACC	GAGACCCCTGCG	TGGTGCAGAA	CGCCAACCCC	GACTGCAAGA	1000
GagProt.ModS	951	CTGGATGACC	GAGACCCCTGCG	TGGTGCAGAA	CGCCAACCCC	GACTGCAAGA	1000
Gag.ModSF2	951	CTGGATGACC	GAGACCCCTGCG	TGGTGCAGAA	CGCCAACCCC	GACTGCAAGA	1000
GagPol.ModSF	1001	CCATCTGAA	GGCTCTCGGC	CCCGCGGCCA	CCCTGGAGGA	GATGATGACC	1050
GagProt.ModS	1001	CCATCTGAA	GGCTCTCGGC	CCCGCGGCCA	CCCTGGAGGA	GATGATGACC	1050
Gag.ModSF2	1001	CCATCTGAA	GGCTCTCGGC	CCCGCGGCCA	CCCTGGAGGA	GATGATGACC	1050
GagPol.ModSF	1051	GCCTGCCAGG	GGGTGGGGGG	CCCCGGCAC	AAGGGCCCGG	TGCTGGCGGA	1100
GagProt.ModS	1051	GCCTGCCAGG	GGGTGGGGGG	CCCCGGCAC	AAGGGCCCGG	TGCTGGCGGA	1100
Gag.ModSF2	1051	GCCTGCCAGG	GGGTGGGGGG	CCCCGGCAC	AAGGGCCCGG	TGCTGGCGGA	1100
GagPol.ModSF	1101	GGCATGAGC	CAGGTGACGA	ACCCGGGAC	CATCATGATG	CAGGGGGCA	1150
GagProt.ModS	1101	GGCATGAGC	CAGGTGACGA	ACCCGGGAC	CATCATGATG	CAGGGGGCA	1150
Gag.ModSF2	1101	GGCATGAGC	CAGGTGACGA	ACCCGGGAC	CATCATGATG	CAGGGGGCA	1150
GagPol.ModSF	1151	ACTTCGGCAA	CCAGGGGAAG	ACCGTCAAGT	GCTTCAACTG	CGGCAAGGAG	1200
GagProt.ModS	1151	ACTTCGGCAA	CCAGGGGAAG	ACCGTCAAGT	GCTTCAACTG	CGGCAAGGAG	1200
Gag.ModSF2	1151	ACTTCGGCAA	CCAGGGGAAG	ACCGTCAAGT	GCTTCAACTG	CGGCAAGGAG	1200
GagPol.ModSF	1201	GGCCACACCG	CCAGGAACGT	CCGGCCCCC	CGCAAGAGG	GCTGCTGGCG	1250
GagProt.ModS	1201	GGCCACACCG	CCAGGAACGT	CCGGCCCCC	CGCAAGAGG	GCTGCTGGCG	1250
Gag.ModSF2	1201	GGCCACACCG	CCAGGAACGT	CCGGCCCCC	CGCAAGAGG	GCTGCTGGCG	1250
GagPol.ModSF	1251	CTGCGGCCGC	GAAGGACACC	AAATGAAAAGA	TTGCACTGAG	AGACAGGCTA	1300
GagProt.ModS	1251	CTGCGGCCGC	GAAGGACACC	AAATGAAAAGA	TTGCACTGAG	AGACAGGCTA	1300
Gag.ModSF2	1251	CTGCGGCCGC	GAAGGACACC	AAATGAAAAGA	TTGCACTGAG	AGACAGGCTA	1300
GagPol.ModSF	1301	ATTTTTTAGG	GAAGATCTGG	CCTTCCTACA	AGGAAAGCC	AGGAATTT	1350
GagProt.ModS	1301	ATTTTTTAGG	GAAGATCTGG	CCTTCCTACA	AGGAAAGCC	AGGAATTT	1350
Gag.ModSF2	1301	ATTTCCCTGGG	CAAGATCTGG	CCCAAGCTACA	AGGGCCGCC	CGGCAAACCTC	1350

GagPol.ModSF	1351	CTTCAGAGCA	GACCGAGGCC	AACAGCCCCA	CCAGAACAGA	GCTTCAGGTT	1400
GagProt.Mods	1351	CTTCAGAGCA	GACCGAGGCC	AACAGCCCCA	CCAGAACAGA	GCTTCAGGTT	1400
Gag.ModSF2	1351	CTGCAGAGCC	GCCCCGAGCC	CACCGCCCCC	CCCGAGGGAA	GCTTCAGGTT	1400
GagPol.ModSF	1401	TGGGGAGGAG	AAAACAACCTC	CCTCTCAGAA	GCAGGAGGCC	ATAGAACAGG	1450
GagProt.ModS	1401	TGGGGAGGAG	AAAACAACCTC	CCTCTCAGAA	GCAGGAGGCC	ATAGAACAGG	1450
Gag.ModSF2	1401	CGGGGAGGAG	AAGACCACCC	CCAGCCAGAA	GCAGGAGCCC	ATCGAACAGG	1450
GagPol.ModSF	1451	AACTGTATCC	TTAACCTTCC	CTCAGATCAC	TCTTTGGCAA	CGACCCCTCG	1500
GagProt.Mods	1451	AACTGTATCC	TTAACCTTCC	CTCAGATCAC	TCTTTGGCAA	CGACCCCTCG	1500
Gag.ModSF2	1451	AGCTGTACCC	CCTGACCAGC	CTGGCGAGCC	TGTTGGCAA	CGACCCCAAGC	1500
GagPol.ModSF	1501	TCACAGTAAG	GATCGGGGGC	CAGCTCAAGG	AGGGCCTGCT	CGACACCGGC	1550
GagProt.Mods	1501	TCACAGTAAG	GATCGGGGGC	CAGCTCAAGG	AGGGCCTGCT	CGACACCGGC	1550
Gag.ModSF2	1501	AGCCAGTAA.	1550
GagPol.ModSF	1551	GCCGACGACA	CGCTGCTGGA	GGAGATGAAC	CTGCCCCGGCA	AGTGGAAAGCC	1600
GagProt.Mods	1551	GCCGACGACA	CGCTGCTGGA	GGAGATGAAC	CTGCCCCGGCA	AGTGGAAAGCC	1600
Gag.ModSF2	1551	1600
GagPol.ModSF	1601	CAAGATGATC	GGGGGGATCG	GGGGCTTCAT	CAAGGGTGGCG	CAGTACGACC	1650
GagProt.Mods	1601	CAAGATGATC	GGGGGGATCG	GGGGCTTCAT	CAAGGGTGGCG	CAGTACGACC	1650
Gag.ModSF2	1601	1650
GagPol.ModSF	1651	AGATCCCCGT	GGAGATCTGC	GGCCACAAAGG	CCATCGGGCAC	CAGTGGTGGTG	1700
GagProt.Mods	1651	AGATCCCCGT	GGAGATCTGC	GGCCACAAAGG	CCATCGGGCAC	CAGTGGTGGTG	1700
Gag.ModSF2	1651	1700
GagPol.ModSF	1701	GGCCCCACCC	CCGTGAACAT	CATCGGGCGC	AACCTGCTGA	CCCGAGATCGG	1750
GagProt.ModS	1701	GGCCCCACCC	CCGTGAACAT	CATCGGGCGC	AACCTGCTGA	CCCGAGATCGG	1750
Gag.ModSF2	1701	1750
GagPol.ModSF	1751	CTGCACCCCTG	AACCTCCCCA	TCAGCCCCAT	CGAGACGGGT	CCCGTGAAGC	1800
GagProt.ModS	1751	CTGCACCCCTG	AACCTCCCCA	TCAGCCCCAT	CGAGACGGGT	CCCGTGAAGC	1800
Gag.ModSF2	1751	1800

FIG. 7D

10 / 131

GagPol.ModSF	1801	TGAAGCCGGG	GATGGACGGC	CCCAAGGTCA	AGCAGTGGCC	CCTGACCGAG	1850
GagProt.ModS	1801	TGAAGCCGGG	GATGGACGGC	CCCAAGGTCA	AGCAGTGGCC	CCTGTAA...	1850
Gag.ModSF2	1801	1850
GagProt.ModS	1851	GAGAAGATCA	AGGCCCTGGT	GGAGATCTGC	ACCGAGATGG	AGAACGGGG	1900
Gag.ModSF2	1851	1900
GagPol.ModSF	1901	CAAGATCAGC	AAGATCGGGC	CCGAGAACCC	CTACAACACC	CCCGTGTTCG	1950
GagProt.ModS	1901	1900
Gag.ModSF2	1901	1900
GagPol.ModSF	1951	CCATCAAGAA	GAAGGACAGC	ACCAAGTGGC	GCAAGCTGCT	GGACTTCGGC	2000
GagProt.ModS	1951	2000
Gag.ModSF2	1951	2000
GagPol.ModSF	2001	GAGCTAACAA	AGGGCACCCA	GGACTCTGG	GAGGTGCAGC	TGGGCATCCC	2050
GagProt.ModS	2001	2050
Gag.ModSF2	2001	2050
GagPol.ModSF	2051	CCACCCCGCC	GGCCTGAAA	AGAACGAGAG	CGTGACCGTG	CTGGACGTGG	2100
GagProt.ModS	2051	2100
Gag.ModSF2	2051	2100
GagPol.ModSF	2101	GCGACGCCCA	CTTCAGGGTG	CCCTGGACA	AGGACTTCG	CAAGTACACC	2150
GagProt.ModS	2101	2150
Gag.ModSF2	2101	2150
GagPol.ModSF	2151	GCCTTCACCA	TCCCCAGCAT	CAACAAAGAG	ACCCCCGGCA	TCCCGTACCA	2200
GagProt.ModS	2151	2200
Gag.ModSF2	2151	2200
GagPol.ModSF	2201	GTACAAACGTG	CTGCCCCAGG	GCTGGAAGGG	CAGCCCCGGC	ATCTTCCAGA	2250
GagProt.ModS	2201	2250
Gag.ModSF2	2201	2250

FIG. 7E

11 / 131

GagPol.ModSF	2251	GCAGCATGAC	CAAGATCCTG	GAGCCCTTCC	GCAAGCAGAA	CCCCGACATC	2300
GagProt.ModS	2251	2300
Gag.ModSF2	2251	2300
GagPol.ModSF	2301	GTGATCTACC	AGTACATGGA	CGACCTGTAC	GTGGCAGCG	ACCTGGAGAT	2350
GagProt.ModS	2301	2350
Gag.ModSF2	2301	2350
GagPol.ModSF	2351	CGGCCAGCAC	CGCACCAAGA	TCGAGGAGCT	GCGCCAGCAC	CTGCTGCGCT	2400
GagProt.ModS	2351	2400
Gag.ModSF2	2351	2400
GagPol.ModSF	2401	GGGGCTTCAC	CACCCCCGAC	AAGAGCACC	AGAAGGAGCC	CCCCTTCCTG	2450
GagProt.ModS	2401	2450
Gag.ModSF2	2401	2450
GagPol.ModSF	2451	TGGATGGCT	ACGAGCTGCA	CCCCGACAAG	TGGACCGTGC	AGCCCCATCAT	2500
GagProt.ModS	2451	2500
Gag.ModSF2	2451	2500
GagPol.ModSF	2501	GCTGCCGAG	AAGGACAGCT	GGACCGTGAA	CGACATCCAG	AAGCTGGTGG	2550
GagProt.ModS	2501	2550
Gag.ModSF2	2501	2550
GagPol.ModSF	2551	GCAAGCTGAA	CTGGGCCAGC	CAGATCTACG	CCGGCATCAA	GGTGAAGCAG	2600
GagProt.ModS	2551	2600
Gag.ModSF2	2551	2600
GagPol.ModSF	2601	CTGTGCAAGC	TGCTGCGGG	CACCAAGGCC	CTGACCGAGG	TGATCCCCCT	2650
GagProt.ModS	2601	2650
Gag.ModSF2	2601	2650
GagPol.ModSF	2651	GACCGAGGAG	GGCCGAGCTGG	AGCTGGCCGA	GAACCCGGAG	ATCCCTGAAGG	2700
GagProt.ModS	2651	2700
Gag.ModSF2	2651	2700

FIG. 7F

12 / 131

GagPol.ModSF	2701	AGCCCGTGCA	CGAGGTGTAC	TACGACCCCA	GCAGGGACCT	GGTGGCCGAG	2750
GagProt.Mods	2701	2750
Gag.ModSF2	2701	2750
GagProt.ModSF	2751	ATCCAGAAC	AGGCCAGGG	CCAGTGGACC	TACCATCT ACCAGAGCC	2800	2800
GagProt.Mods	2751	2800
Gag.ModSF2	2751	2800
GagPol.ModSF	2801	CTTCAGAAC	CTGAAGACCG	GCAAGTACGC	CCGCATGCGC	GGGCCACACA	2850
GagProt.Mods	2801	2850
Gag.ModSF2	2801	2850
GagPol.ModSF	2851	CCAACGACT	GAAGCAGCTG	ACCGAGGCCG	TGCAAGGGT	GAGCACCCGAG	2900
GagProt.Mods	2851	2900
Gag.ModSF2	2851	2900
GagPol.ModSF	2901	AGCATCGTGA	TCTGGGCAA	GATCCCAAG	TTCAAAGCTGC	CCATCCAGAA	2950
GagProt.Mods	2901	2950
Gag.ModSF2	2901	2950
GagPol.ModSF	2951	GGAGACCTGG	GAGGCCTGGT	GGATGGAGTA	CTGGCAGGGC	ACCTGGATCC	3000
GagProt.Mods	2951	3000
Gag.ModSF2	2951	3000
GagPol.ModSF	3001	CCGAGTGGGA	GTTCGTGAAC	ACCCCCCCC	TGGTAAGCT	GTGGTACCAAG	3050
GagProt.Mods	3001	3050
Gag.ModSF2	3001	3050
GagPol.ModSF	3051	CTGGAGAAC	AGCCCCATGT	GGGGCCGAG	ACCTTCTACG	TGGACGGCGC	3100
GagProt.Mods	3051	3100
Gag.ModSF2	3051	3100
GagPol.ModSF	3101	CGCCAACCGC	GAGACCAAGC	TGGCCAAGGC	CGGCTACGTG	ACCGACCCGGC	3150
GagProt.Mods	3101	3150
Gag.ModSF2	3101	3150

FIG. 7G

13 / 131

GagPol.ModSF	3151	GCCGCCAGAA	GGTGGTGAGC	ATCGCCGACA	CCACCAACCA	GAAGACCGAG	3200
GagProt.Mods	3151	3200
Gag.ModSF2	3151	3200
GagProt.Mods	3201	3210	3220	3230	3240	3250	3250
GagProt.Mods	3201	CTGGAGGCCA	TCCACCTGGC	CCTGCAGGAC	ACGGGCCTGG	AGGTGAACAT	3250
Gag.ModSF2	3201	3250
GagProt.Mods	3251	3260	3270	3280	3290	3300	3300
GagProt.Mods	3251	CGTGACCGAC	AGCCAGTAGG	CCCTGGCAT	CATCCAGGGC	CAGCCCCGACA	3300
Gag.ModSF2	3251	3300
GagProt.Mods	3301	3310	3320	3330	3340	3350	3350
GagProt.Mods	3301	AGAGGGAGAG	CGAGCTGGTG	AGCCAGATCA	TGAGGCAGCT	GATCAGAG	3350
Gag.ModSF2	3301	3350
GagProt.Mods	3351	3360	3370	3380	3390	3400	3400
GagProt.Mods	3351	GAGAAGGGTGT	ACCTGGCTTG	GGTGGCCGCC	CACAAAGGGAA	TGGGGGCCAA	3400
Gag.ModSF2	3351	3400
GagProt.Mods	3401	3410	3420	3430	3440	3450	3450
GagProt.Mods	3401	CGAGCAGGTG	GACAAGGCTGG	TGAGGCCCGG	CATCCGCAAG	GTGCTGTTC	3450
Gag.ModSF2	3401	3450
GagProt.Mods	3451	3460	3470	3480	3490	3500	3500
GagProt.Mods	3451	TGAACCGGCAT	CGACAAAGGCC	CAGGAGGAGC	ACGAGAAAGTA	CCACAGCAAC	3500
Gag.ModSF2	3451	3500
GagProt.Mods	3501	3510	3520	3530	3540	3550	3550
GagProt.Mods	3501	TGGCGGGCCA	TGGCCAGCGA	CTTCAACCTG	CCCCCGTGG	TGGCCAAGGA	3550
Gag.ModSF2	3501	3550
GagProt.Mods	3551	3560	3570	3580	3590	3600	3600
GagProt.Mods	3551	GATCGTGGCC	AGCTGCGACA	AGTGCCAGCT	GAAGGGCGAG	GCCATGCACG	3600
Gag.ModSF2	3551	3600

FIG. 7H

14 / 131

GagPol.ModSF	3601	GCCAGGTGGA	CTGCAGCCCC	GGCATCTGGC	AGCTGGACTG	CACCCACCTG	3650
GagProt.ModS	3601	3650
Gag.ModSF2	3601	3650
GagPol.ModSF	3651	GAGGCCAAGA	TCAATCCTGGT	GGCCGTGCAC	GTGGCCAGCG	GCTACATCGA	3700
GagProt.ModS	3651	3700
Gag.ModSF2	3651	3700
GagPol.ModSF	3701	GGCCGAGGTG	ATCCCCGGCCG	AGACC GGCCA	GGAGACC CGCC	TACT CCTG C	3750
GagProt.ModS	3701	3750
Gag.ModSF2	3701	3750
GagPol.ModSF	3751	TGAAGGCTGGC	CGGGCCGGCTGG	CCCGTGAAAGA	CCATCCACAC	CGACAA CGGC	3800
GagProt.ModS	3751	3800
Gag.ModSF2	3751	3800
GagPol.ModSF	3801	AGCAACTTCA	CCAGCACAC	CGTGAAGGCC	GCCTGCTGCT	GGGCCGGCAT	3850
GagProt.ModS	3801	3850
Gag.ModSF2	3801	3850
GagPol.ModSF	3851	CAA GAGGGAG	TTCGGCATCC	CCTACAACCC	CCAGAGCCAG	GGCCGGTTGG	3900
GagProt.ModS	3851	3900
Gag.ModSF2	3851	3900
GagPol.ModSF	3901	AGAGCATGAA	CAACGAGGTG	AAGAGATCA	TCGGCCAGGT	GCGGACCAAG	3950
GagProt.ModS	3901	3950
Gag.ModSF2	3901	3950
GagPol.ModSF	3951	GCCGAGCACC	TGAAGACCC	CGTGCAGATG	GCCGTGTTC	TCCACAACTT	4000
GagProt.ModS	3951	4000
Gag.ModSF2	3951	4000
GagPol.ModSF	4001	CAAGGGCAAG	GGGGGCATCG	GGGGCTACAG	CGCCGGGAG	CGCATCGTGG	4050
GagProt.ModS	4001	4050
Gag.ModSF2	4001	4050

FIG. 71

FIG. 7J

GagPol.ModSF	4051	ACATCATCGC	CACCGACATC	CAGACCAAGG	AGCTGCAGAA	GCAGATCACC	4100
GagProt.ModS	4051	4100
Gag.ModSF2	4051	4100
GagPol.ModSF	4101	AAGATCCAGA	ACTTCCGGGT	GTACTACCGC	GACAACAGG	ACCCCCCTGTG	4150
GagProt.ModS	4101	4150
Gag.ModSF2	4101	4150
GagPol.ModSF	4151	GAAGGGCCCC	GCCAAAGCTGC	TGTGGAAGGG	CGAGGGGCC	GTGGTGATCC	4200
GagProt.ModS	4151	4200
Gag.ModSF2	4151	4200
GagPol.ModSF	4201	AGGACAAACAG	GACATCAAG	GTGGTGCCTC	GCCGCAAGGC	CAAGATCATC	4250
GagProt.ModS	4201	4250
Gag.ModSF2	4201	4250
GagPol.ModSF	4251	CGCGACTACG	GCAAGCAGAT	GGCGGGGAC	SACTGCGTGG	CCAGCCGCCA	4300
GagProt.ModS	4251	4300
Gag.ModSF2	4251	4300
GagPol.ModSF	4301	GGACGAGGAC	TAG	4350
GagProt.ModS	4301	4350
Gag.ModSF2	4301	4350

16 / 131

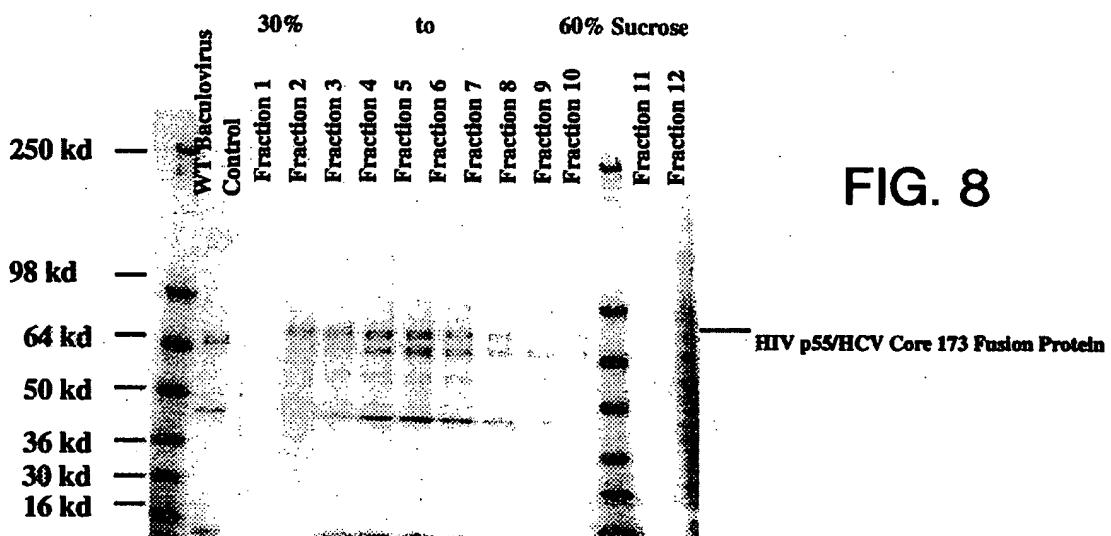


FIG. 8

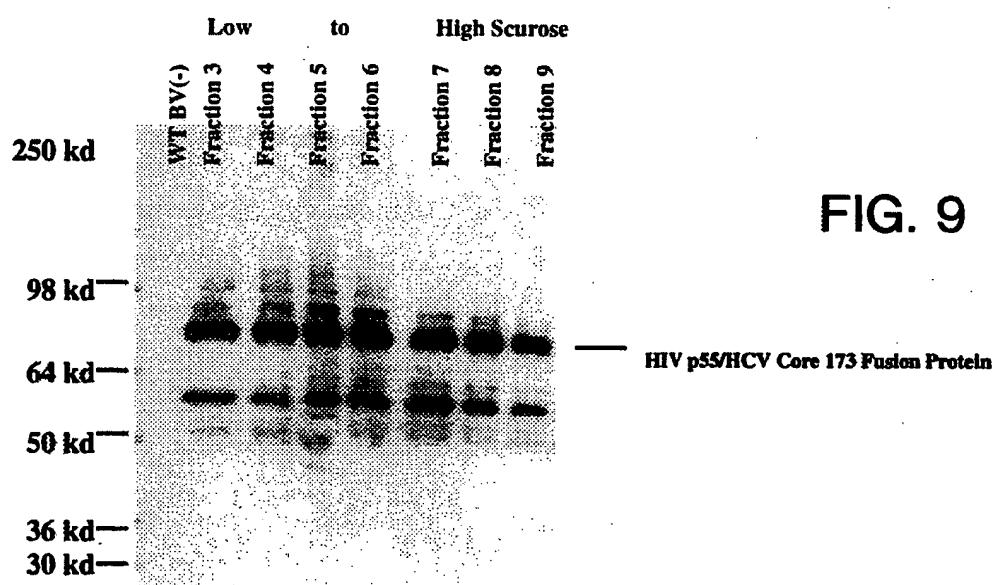
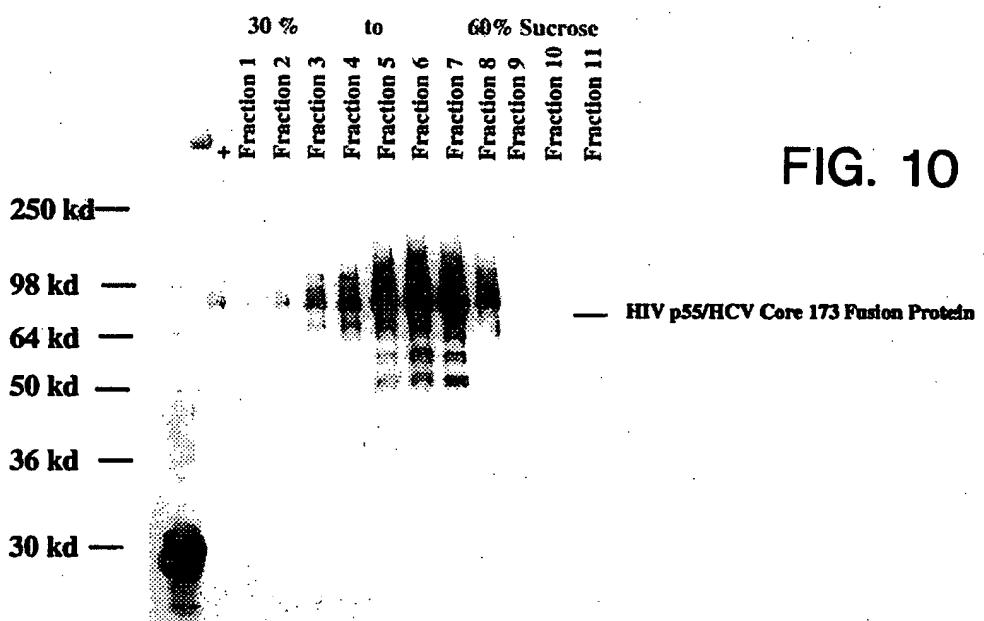


FIG. 9



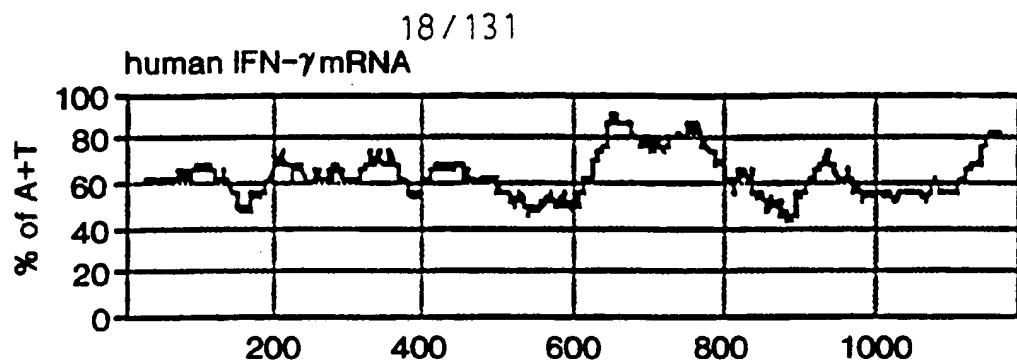


FIG. 11A

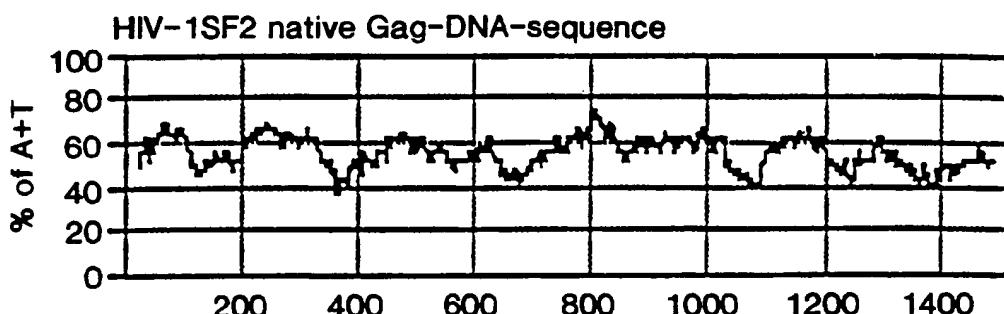


FIG. 11B

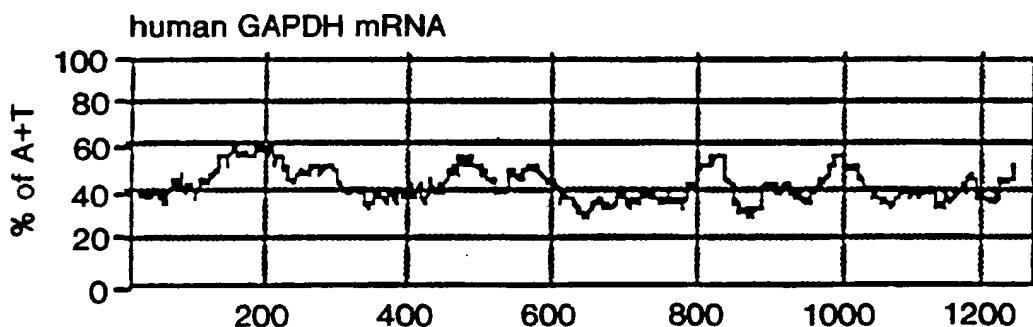


FIG. 11C

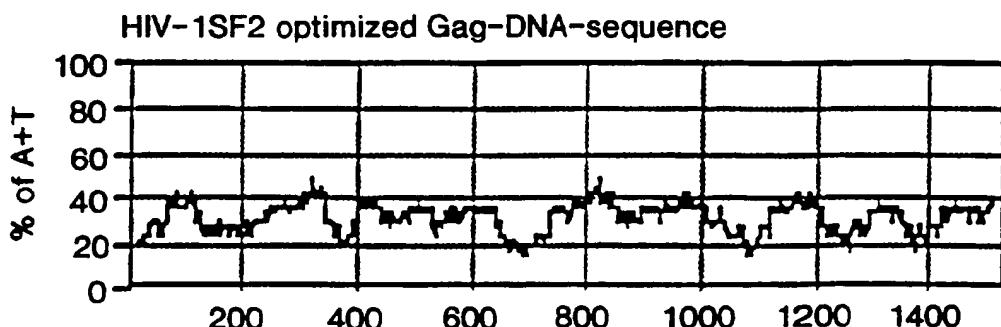


FIG. 11D

native HIV-1SF2 gag-polymerase

ATGGGTGCCGAGGGCTCGTATTAAAGCGGGGGAGAATTAGATAAATGGAAAAAATTGGTTAAGGCCAGGGAAAAG

Inact. 1

A	T	A	A	T	A	G	T	T	A	A	C	A	T	A	T	G	G	C	C	C	C
G	G	C	C	G	C	G	C	C	G	C	G	C	C	G	C	G	C	G	C	G	C

Inact. 2

A	C	A	G	A	G	C	A	A	T	G	T	T	A	T	G	G	C	A	C	A	C	A
T	T	A	T	A	T	G	T	A	T	G	A	G	T	A	T	G	C	A	C	A	G	A

Inact. 3

A	G	G	A	A	G	G	A	A	T	G	T	G	C	A	G	G	C	A	G	C	A	G
G	C	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C

Inact. 4

G	A	G	A	G	C	A	A	T	G	T	G	C	A	G	G	C	A	G	C	A	G	C
T	T	C	C	G	G	C	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G

AGC

AGCAAAATTACCCATAGTGCAGAACCTACAGGGCAAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCA

TGGT

AGTAGTAGAGAGAGAGGGCTTCAGCCCAGAAGTAATACCCATGTTTCAGCATATCAGAAGGGGCCACC

Inact. 5

C	C	A	G	A	T	T	A	A	C	A	G	G	C	A	T	G	T	A	A	G	A	G
G	G	C	C	G	T	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G

CCAC

AGATTAAACACCATGCTAACACAGTGGGGCATCAAGCAGCCATGCAAATGTTAAAAGAGACTATCAATGAGAAGG

GAGGAAG

GACTACTAAGTACAGTACCTACGAAATGGATGGATGACAATAATCCACCTATCCCAGTA

Inact. 6

G	G	A	A	T	T	A	A	T	A	T	G	T	A	G	A	T	T	C	T	G	G	A
G	G	C	G	G	T	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G

ATAAGACAAGGACCAAGGAACCCCTTTAGAGATTATGTAGACCGGTCTATAAAACTCTAAGGCCAAGCTCA

FIG. 12A

20 / 131

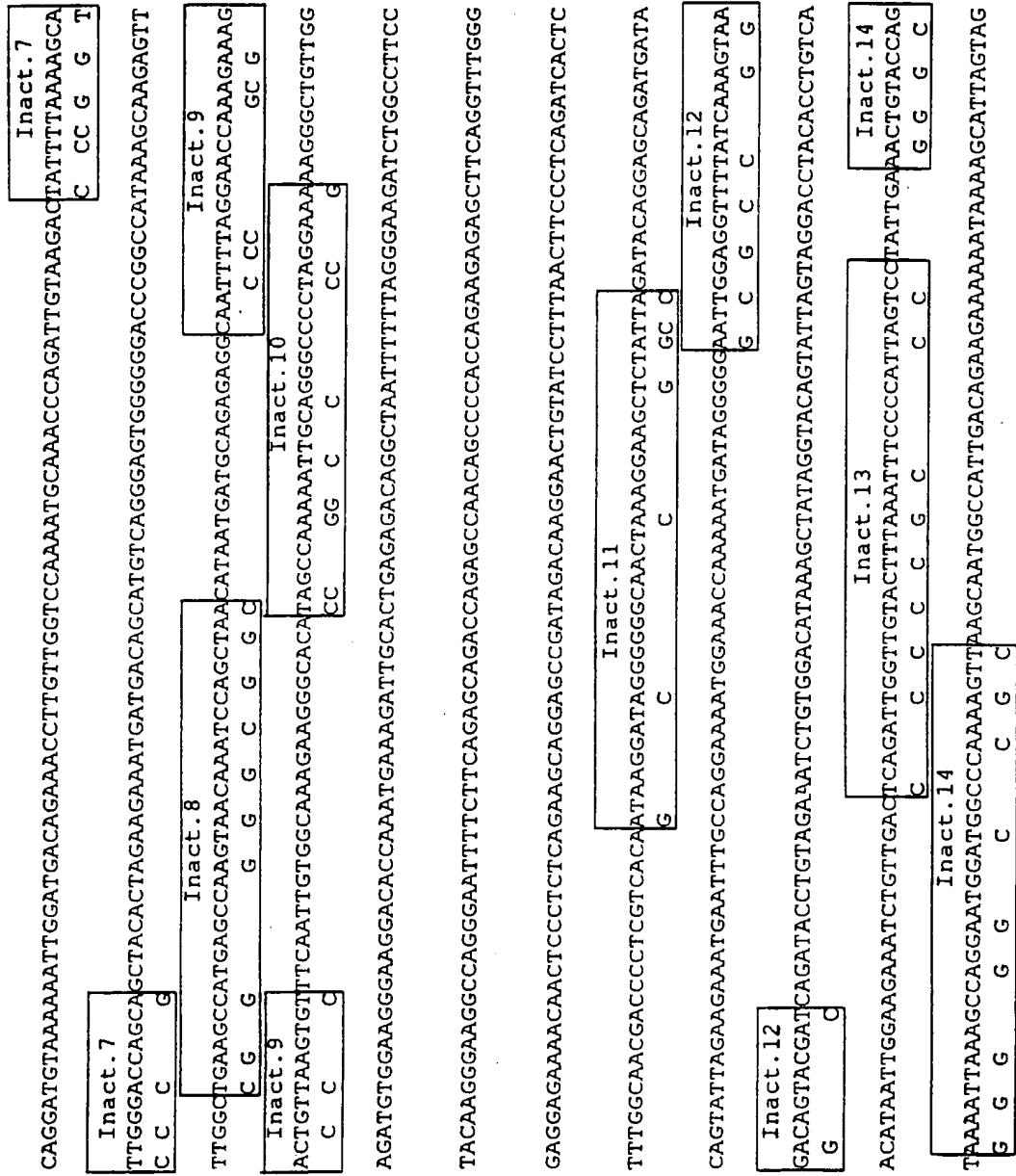
**FIG. 12B**

FIG. 12C

AGATATGTACAGAAATGGAAAAGGAAATTCAAAAAATTGGGCCTGAAAATCCATACAATACTCCAGTATTTG
 CTATAAAGAAAAAGACAGTACTAATGGAAAAACTAGTAGATTCAAGAACCTTAATAAAAGAACTCAAGACTCT
 GGGAAAGTTCAAGTTAGGAATACCAACCCGGAGGGTTAAAAAGRAAAAATCAGTAACAGTATTGGATGGGGTATG
 CATACTTTCAAGTCAGTTCCCTTAGATAAAGACTTAAAGTAGAAAGTATACTGCATTACCATACCTAGTATAAAATGAGAAC
 CAGGGATTAGATATCAGTACAATGTGCTGCCACAGGGATGGAAAAGGATCACCAGCAATATCCAAGTAGCATGACA
 AAATCTTAGAGCCCTTTAGAAAAACAGAATCCAGACATAGTTATCTATAATACATGGATGATTGTAGGATCTG
 ACTTAGAAATAGGGCACATAGAACAAAAATAGAGGAACACTGAGACAGCATTGTGAGGTGGGATTACACACAG
 ACAAAACATCAGAANGAACCTCATTCCCTGGATGGTTATGAACTCCATCTGTATAAATGGACAGTACAGCTA
 TAATGCTGCCAGAAAANGACAGCTGGACTGTCATGACATACAGAAGTTAGTGTAAACAGTTCATGAACTGTTAGTGGAAAATGAAATTG
 TTATGAGGGATTAAAGTAAGCAGTTATGTAACCTCTAGGATGAACTCCCTAGAGAACCCAAAGCAACTAACAAGTDTATACCACTAA
 CAGAAAGCAGAGCTAGAACCTGGAGAAAACAGGGAGATTCTAAAGAACCCATACATGGAGTAAATGACCAT
 CAAAAGACTTAGTAGCAGAAAATACAGAACGGGGCAAGGGCAAGGCAATGGACATATCAAATTATCAAGGCCATTAA
 ATCTGAAAACAGGAAAGTATGCAAGGGATGAGGGTGCACACTAACTGATGTAACAGTTCATGAACTGTTAGTGGGAGTGC
 AAGTATCCACAGAAAAGCATAGTAATATGGGAAAGATTCTAAATTAAACTACCCATACAAAGGAACATGGGAG
 CATGGGGATGGAGTATGGCAAGCTACCTGGATTCGTTGAGTTGTCATAACCCCCTCCTTAGTGAATTT
 GTTACCCAGTTAGAGAAAAGAACCCATAGTAGGAGCAGAACACTTCTATGTAGATGGGAGCTAATAGGGAGCTAAAT
 TAGGAAAAGCAGGGATATGTTACTGACAGAGGAAACAAAAGTTGTCATAGGTTAGTCAGTCAAATATAGAGCAGTTAATAAAAGGAAA
 AATTACAAAGCAATTCTAGCTTTGCAAGGATTGGGATTAGAACTAGTGAATCAAGGATAATTAGTCAAGTAAAT
 GAATCATTCAAGCACACCAGATAAGAGTGAATCAAGGATAATTAGGAAATTGGGAAATTGAGATAATTAGTCAGTGCTGGAA
 AGGTCTACCTGGCATGGTACAGCACACAAAGGAAATTGGGAAATTGAGATAATTAGGAAATTGGGAAATT
 TCAGGAAAAGTACTATTGTAATGGAAATTAGATAAAGGCCCAAGAACGACATGAGAAAATATCAGCTAAATTGGAGGCAA
 TGGCTAGTGTGATTAAACCTGCCACCTGTAGTAGCAAAAGAAAATAGTAGCCAGCTGTGATAAAATGTCAGCTAAAGGAG
 AAGCCATGCTGATGGACAGTAGCTAGTGTAGCTGCAAGGAAATTGGCAACTAGTGAATTGGCAATCTAGAAGGGAAAATTATCC
 TGTTAGCGATTCTAGTGGCCAGTGGGATATAAGAGCAGGAAATTGGCAATCTAGAAGGAAATTGGCAATTTCAGCAGGGCAGGAAACGGCATATT
 TTCTCTAAATTAGCAGGAAGATGGCCAGTAAAGCAGGAAATTGGCAATTCCCTCAAAATGGCAATTCCCTAAAGTCAGGGTAGTAG
 AATCTATGAAATATGAAATTAAAGGAAATTAGGACAGGTAAGAGTCAAGGAAATTGGCAATTCCCAAGCTTAAGACAGCAGTACAAA
 TGGCAGTATCATCCACAAATTTPAAGAAGGGGGATTGGGGATACAGTGGGGGAAAGGAAATTAGTGAACCTAGTACAGTAA
 TAGCAACAGACATACAAACTAAAGAAACTACAAAGCAAAATTACAAATTCAAAATTTGGGTTTATTCAGGGACA
 ACAAAGATCCCTTTGGAAAGGACCAAGCTCTGGAAAGGTGAAGGGCAGTGTAAATACAAGATAATAGT
 ACATAAAAGTAGTGCACAGGATGGGATTAG
 CAAGTAGACAGGATGGGATTAG

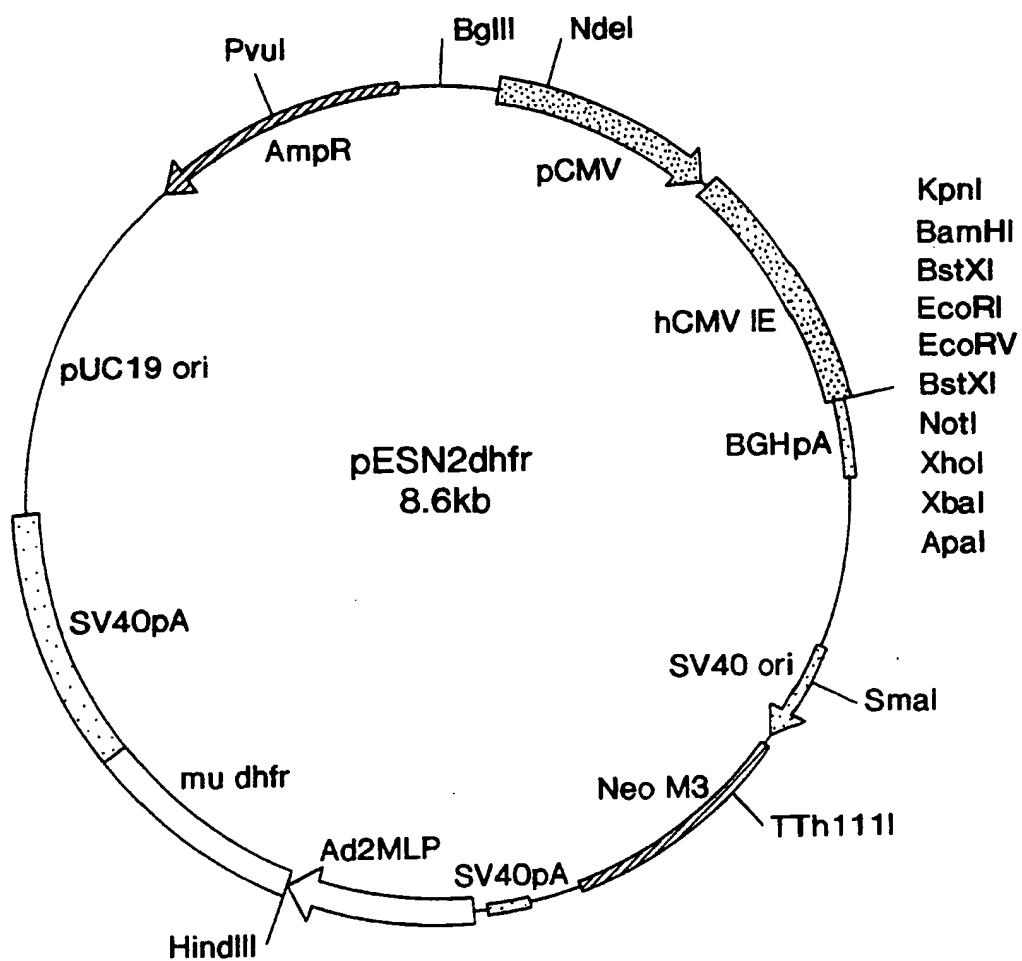


FIG. 13A

23 / 131

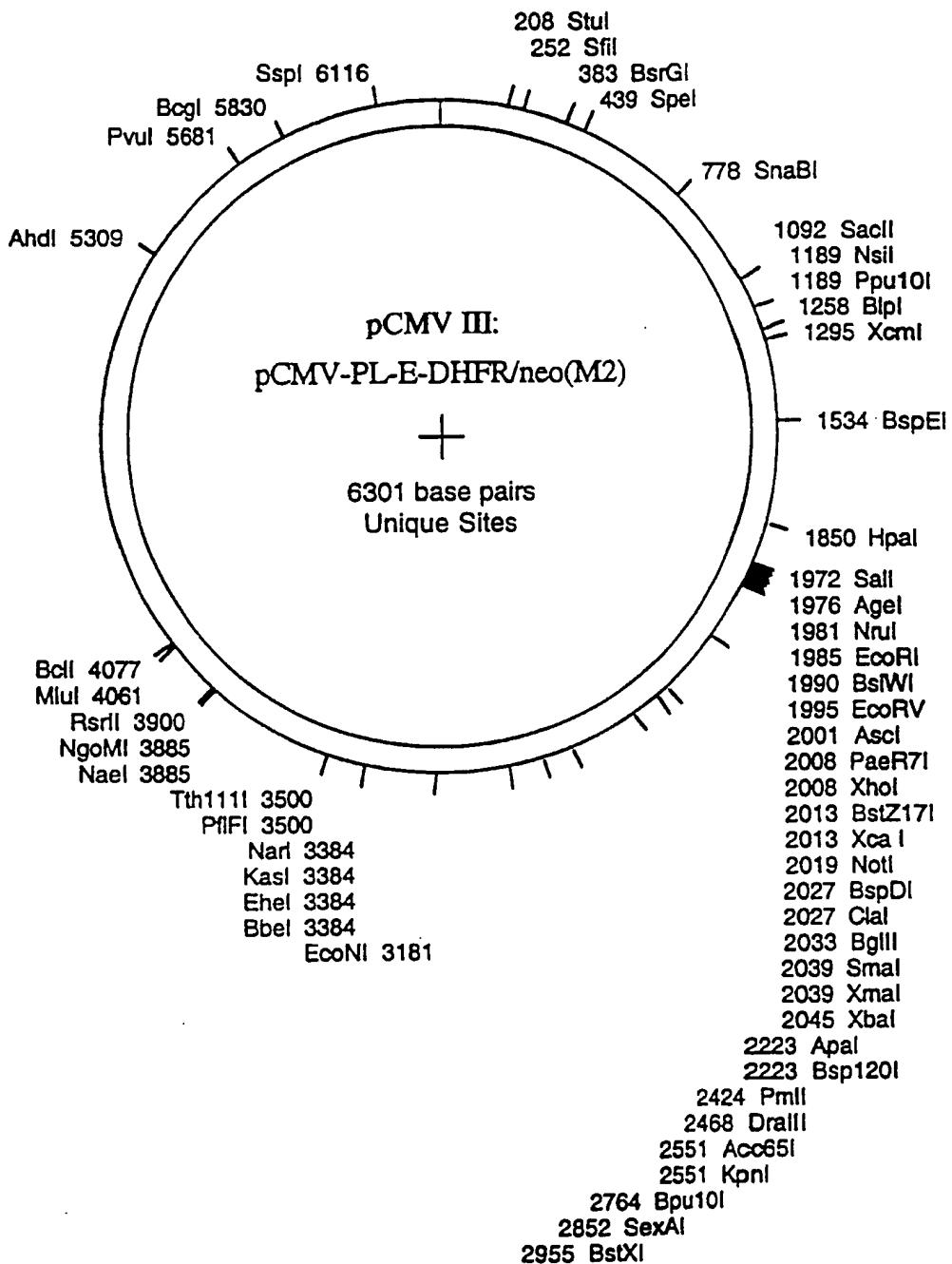


FIG. 13B

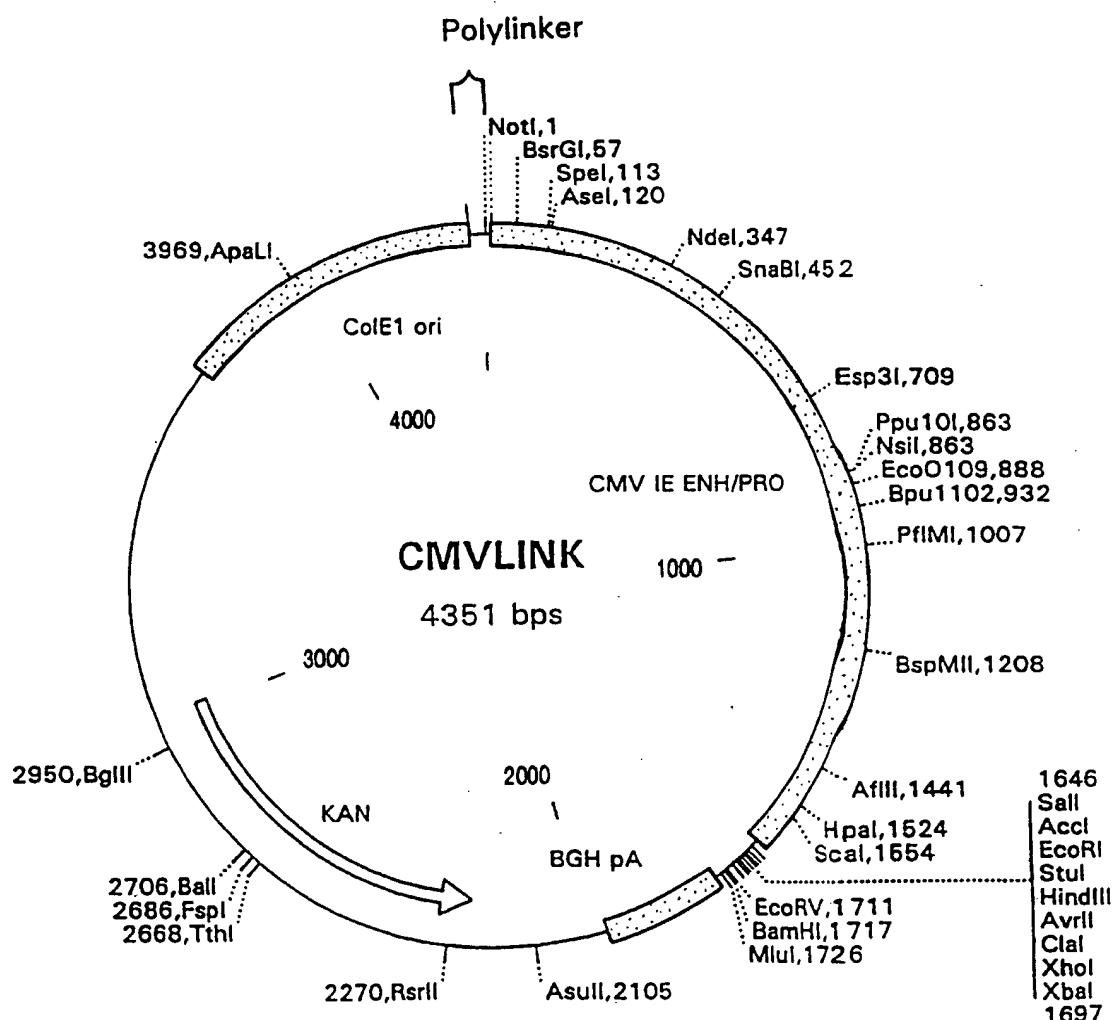


FIG. 14

25 / 131

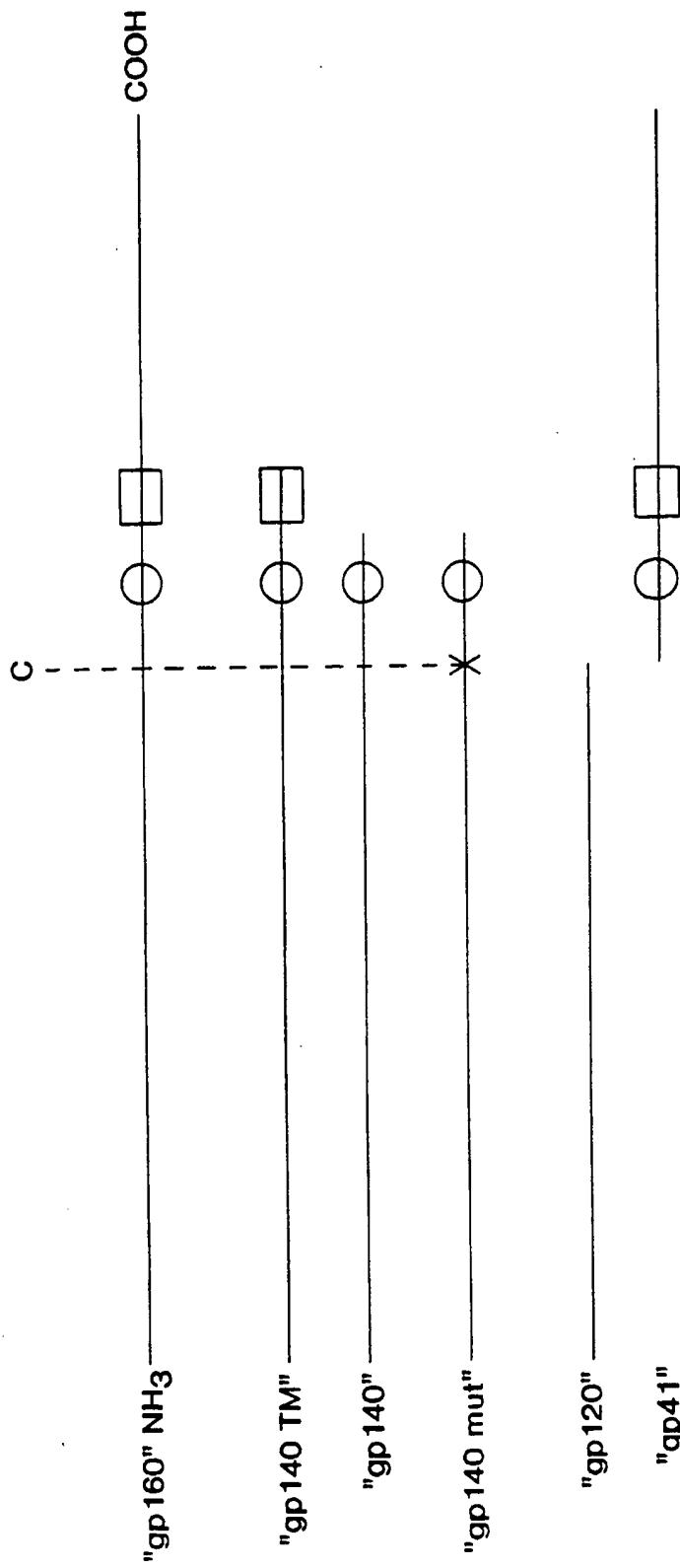


FIG. 15

9P120wtSF162

GTAGAAAAATTGGGGTACAGTCTATTATGGGGTACACTGTTGGAAAGAACCAACCTT
 GTGCATCAGATGCCATTGACACAGGTCATAAGGCTATGACAGGTTACATAATTGCTGGCCACACATGCCCTGTGTAACCCAC
 AGACCCCTAACCCACAAGAAATAGTATTGGAAAATGTGACAGAAAATTAAACATGTGGAAAATAACATGCTAAGGTAAACATG
 GTAGAACAGATGCCATTGAGGATATACTCGTTATGGATCAAAGCTTAAGGTAAAGCTTAAGGTAAAGTAAACCTTAAAGTAAACATG
 CACTCTGTGTTACTCTACATTGCACTAAATTGCTCTTCAAGGTCAACCACAGCATAGAAATTGGAAAGTAGTAAAGTGAAGAAG
 GATGGACAGAGGAGAAATAAAAATTGCTTAAGGTAAATGATAAGGATAATAGATAATGATAATAACAGCTATAATTGATAAA
 GAATATGCACTTTTATAAACCTTGATGTAGTACCAATTGCTTAAGGAAATTGCTTAAGGAAATTGCTTAAGGAAATTGCTTAAGGAA
 ATTGTAACACCTCAGTCATTACACAGGCTGTCCAAGGTATCCTTGAACCAATTCCCATACATTATTG
 TGCCCCGGCTGGGTTGGCATCTAAAGTGTAAATGATAAGGAAGTTCATGGATCAGGACCATGTACAAT
 GTCAAGCACAGTACAATGTACACATGGAAATTAGGCCAGTAGTGTCAACTCAATTGCTGTAAATGCCAGTC
 TAGCAGAAAGGGTAGTAATTAGATCTGAAATTGCTGAAATTTCACAGACAATGCTAAACACTATAATGTAACAGCT
 GAAGGAATTCTGTAGAAATTGCTGAAATTGCTGAAATTTCACAGACAATGCTAAACATAACAGAAAGTATAACTATAGGACCC
 GGGAGAGCATTATTGCAACAGGAGACATAATGGAGATAAAGGAGATAAAGACAAGCACATTGTAACATTAGTGGAG
 AAAATGGATAAACACTTTAACAGATACTGTTACAAAATTCAAGGCCAAATTGGGAATAAAACATACT
 CTTAAGCAATCCTCAGGAGGGACCCAGAAATTGTAATGCACTGTTAATGCTTAATTGGAAGTGGGAATTTTTC
 TACTGTAATTCAACACAGCTTTTAATAGTACTTGGATAAATACTATAAGGCCAAATAACACTATAATGGAA
 CTATCACACTCCCATGCGAAATAAACAAATTAAACAGGTGGCAGGAAGTAGGAAAGCAATGTATGC
 CCCTCCCATCAGGGACAATTAGATGCTCATCAAATTACAGGACTGCTTAAACAGAGATGGTGG
 AAAGAGATCAGTAACACCCAGAGATCTCAGACCTGGAGATAATGGGACATTGGGAAGTG
 AATTATAATAAAGTAGTAAATTGAGCATTAGGAGTAGCCACCAAGGCAAAAGAGAAGAGT
 GGTGCAGAGAGAAAAAGA

FIG. 16
(SEQ ID NO:30)

gp140wtSF1162

GTAGAAAAATTGTGGGTACAGTCTATATGGGGTACCTGTGTGAAAGAACCAACCAACTCTATT
 GTGCATCAGATGCTAACGCCCTATGACACAGGCTACATAATGTCTGGCCACACATGGCTGTGTA
 AGACCCTAACCAAGAAATAGTATTGGAAAATTGTGACAGAAAATTAAACATGTGAAAAATAACATG
 CTAGAACAGATGCTAGGATATAATCAGTTATGGATCAAAGCTAAAGCTAATGGTAAAGTTAACCC
 CACTCTGTGTACTCTACATGGCACTAAATTGCTCTTCAAGGTACCCAAAGCTAAAGTAGTAATTGGA
 GATGGACAGAGGAGAAATAAAATTGGATCAAAGCTAACTGTTAAGGCTAAAGCTAAAGTGCAGAA
 GAATATGCACTTTTATAAACCTTGATGTTAGTACCAATAGATAATGATAATACAAGCTATAATTGATA
 ATTGTAACACCTCAGTCATTACACAGGCCGTCTCAAAAGGTATCCTTGAACCCAATTCCCCATA
 TGCCTCGGCTGGTTTGCAGATCTAAAGTGTAAATGATAAGGTTCAATGGATCAGGACCATGTTAA
 GTCAAGCACAGTACAATGTACACATGGAAATTAGGCCAGTAGTGTCAACTCAATTGCTGAGTC
 TAGCAGAAGAAGGGTAGTAATTAGATCTGAAATTACAGACATAATGCTAAAGCACAATGCTAA
 GAGGAATCTGTAGAAATTAACTGTACAAAGACCTAACAAATAACAGATAATGGAGATA
 GGGAGAGCCATTATGGCAACAGGGAGACATAATGGAGATAAGACAAGCACAATTGTAACATTAGTGGAG
 AAAATGGATAAACACTTTAAACAGATAAGTTACAAAATTACAAGCACAATTGGAAATAACAAATA
 CTTAAGCATTACCTCAGGAGGGGACCCAGAAATTGTAATGCAAGTTTAATTGTGGAGGGAAATTTC
 TACTGTAATTCAACACAGCTTTAAATAGTACTTGGAAATAACTATAGGGCAAAATAACACTAATGGAA
 CTATCACACTCCCATTGCGAAATTAAACAGGTGGCAGGAAGTAGGAAAAGCAATTGTATGC
 CCCCTCCATTAGGGACAATTAGATGCTCATCAAATTACAGGACTGCTATTAAACAAAGAGATGGTGGT
 AAGAGAGATCAGTAACACCACCGAGATCTTCAGACCTGGAGGTGGAGATAAGGGACAATTGGAGAAAGT
 AATTATATAAAGTGTAAATTGAGCATTAGGAGTAGCACCACCAAGGAAAGAGAAAGAGT
 GGTGCAAGAGAAAAGAGCAGTGACGGCTAGGAGCTATGTTCCCTGGGTTCTGGGTTCTGGTATA
 ACTATGGGGCAGGGTCACTGACGGTACAGGGCAAGACAATTGCTGGTAAACTCACAGCTGGGCATCAAGCA
 AGAACAAATTGCTGAGAGCTATTGAGGGCAGAACAGCATCTGTGGGAAAGGATCAACAGCTCCTAGGG
 GCTCCAGGGCAAGAGTCTGGCTGGAAAGATAACCTAAAGGATCAACAGCTCCTAGGGATTGGTTGC
 TCTGGAAAAACTCATTGCAACCAACTGGCTGGGTTGGAGTAGTGGAGTAAATCTGGATCAGA
 TTGGAAATTAAACATGCAACCTGGGTGGAGTAGTGGAGTAAATGACAATTACACAAACTTAATACACCT
 AATGGAAGAATCGCAGAACAAAGAAATTAGAATTAGAATTGGGATAAGTGGCAAGT
 TTCTGGGAATTGGTTGACATATCAGGCTGGGTATATA

FIG. 17
(SEQ ID NO:31)

28 / 131

gp160wtSF162

GTAGAAAAATTGGGGTCACAGTCTATTATGGGGTACCTGTGGAAAGAAGCAACCACCACTCTATTT
 GTGCATCAGATGCTAACGCCTATGACACAGAGGTACATAATGTCTGGGCCACACATGCCGTGTACCCAC
 AGACCCCTAACCCACAAGAAATAGTATTGGAAAATGTGACAGAAAATTAAACATGTGGAAAATAACATG
 GTAGAACAGATGCATGAGGATATAATCAGTTATGGGATCAAAGTCTAAAGCATGTGAAAGTTAACCC
 CACTCTGTGTTACTCTACATTGCACTAATTGAAGAATGCTACTAACCAAGAGTAGTAATTGGAAAGA
 GATGGACAGAGGAGAAATAAAAATTGCTCTTCAAGGTACCCACAAGCATAAGAAAATAAGATGCAGAAA
 GAATATGCACTTTTATAAACCTTGATGTAGTACCAATAGATAATGATAATACAAGCTATAAATTGATAA
 ATTGTAACACCTCAGTCATTACACAGGCTGTCCAAGGTATCCTTGAACCAATTCCCACATCATTATTG
 TGCCCCGGCTGGTTTGCATTCTAAAGTGTAAATGATAAGAAGTCAATGGATCAGGACCATGTACAAAT
 GTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTGTCAACTCAATTGCTGTTAAATGGCAGTC
 TAGCAGAAGAAGGGGTAGTAATTAGATCTGAAAATTTCACAGACAATGCTAAACTATAATAGTACAGCT
 GAAGGAATCTGTAGAAATTAAATTGTACAAGACCTAACATAATACAAGAAAAGTATAACTATAGGACCG
 GGGAGAGCATTATGCAACAGGAGACATAATAGGAGATATAAGACAAGCACATTGTAACATTAGTGGAG
 AAAATGGAATAACACTTAAACAGATAGTTACAAAATTACAAGCACAATTGGGATAAAAACAATAGT
 CTTTAAGCAATCCTCAGGAGGGACCCAGAAATTGTAATGCACAGTTTAAATTGTGGAGGGGAAATTTC
 TACTGTAATTCAACACAGCTTTAATAGTACTTGGAAATAATACTATAGGGCAAATAACACTAATGGAA
 CTATCACACTCCATGCAGAATAAAACAAATTATAACAGGGCAGGAAGTAGGAAAAGCAATGTATGC
 CCCTCCCATCAGAGGACAAATTAGATGTCATCAAATATTACAGGACTGCTATTAAACAAGAGATGGTGGT
 AAAGAGATCAGTAACACCACCGAGATCTCAGACCTGGAGGTGGAGATATGAGGGACAATTGGAGAAGTG
 AATTATATAAAATAAAGTAGTAAAATTGAGCCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGT
 GGTGCAGAGAGAAAAAGAGCAGTGACGCTAGGAGCTATGTTCTTGGGTCTGGGAGCAGCAGGAAGC
 ACTATGGCGCAGGGCACTGACGCTGACGGTACAGGCCAGACAATTATTGCTGGTATAGTGCACAGC
 AGAACAAATTGCTGAGAGCTATTGAGGCGAACAGCATCTGTTGCAACTCACAGTCTGGGCATCAAGCA
 GCTCCAGGCAAGAGTCTGGGTGTGGAAAGATAACCTAAAGGATCAACAGCTCTAGGGATTGGGTG
 TCTGGAAAACTCATTGCAACACTGCTGTGCTTGGAAATGCTAGTTGGAGTAATAAATCTCTGGATCAGA
 TTTGGAAATAACATGACCTGGATGGAGTGGAGAGAGAAATTGACAATTACACAAACTTAATACACCTT
 AATTGAAGAATCGCAGAACCAACAAGAAAAGAATGAACAAGAATTATTAGAATTGGATAAGTGGCAAGT
 TTGTGGAATTGGTTGACATATCAAATGGCTGTGGTATATAAAATTCTATAATGATAGTAGGAGGTT
 TAGTAGGTTAAGGATAGTTTACTGTGCTTCTATAGTGAATAGAGTTAGGCAGGGATACTCACCATT
 ATCATTTCAGACCCGCTTCCCAGCCCCAAGGGGACCCGACAGGCCAGGAATCGAAGAAGAAGGTGGA
 GAGAGAGACAGAGACAGATCCAGTCCATTAGTGCATGGATTATTAGCACTCATCTGGGACGATCTACGGA
 GCCTGTGCCTCTTCAGCTACACCGCTTGAGAGACTTAATCTGATTGCAGCGAGGATTGTGAAACTTCT
 GGGACGCAGGGGTGGGAAGCCCTCAAGTATTGGGGAAATCTCTGCAGTATTGGATTCAAGGAACATAAG
 AATAGTGTGTTAGTTGATGCCATAGCTATAGCAGTAGCTGAGGGACAGATAGGATTATAGAAG
 TAGCACAAAGAATTGGTAGAGCTTTCTCCACATACCTAGAAGAATAAGACAGGGCTTGAAAGGGCTT
 GCTATAA

FIG. 18

(SEQ ID NO:32)

gp120.modSF162

FIG. 19
(SEQ ID NO:33)

gp120.mods1162.de1v2

FIG. 20
(SEQ ID NO:34)

gp120.modsP162.delV1V2

FIG. 21
(SEQ ID NO:35)

32 / 131

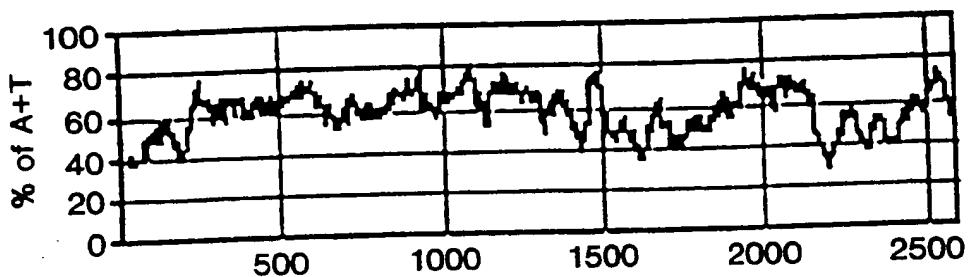


FIG. 22A

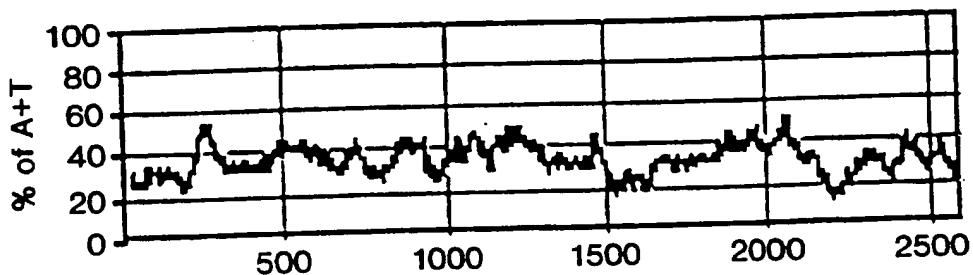


FIG. 22B

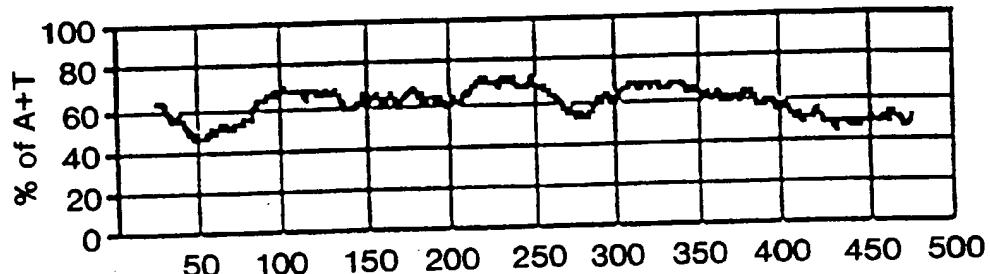


FIG. 22C

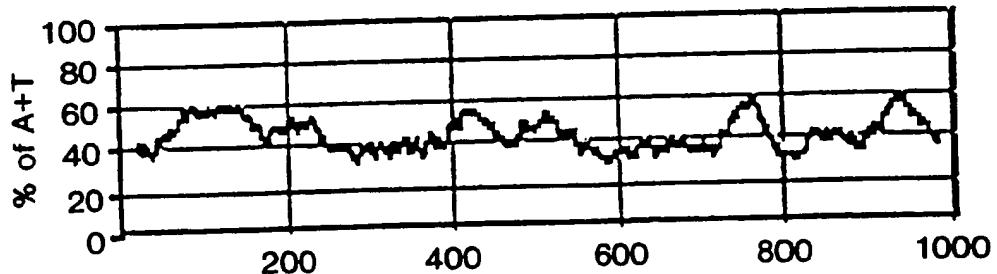


FIG. 22D

33 / 131

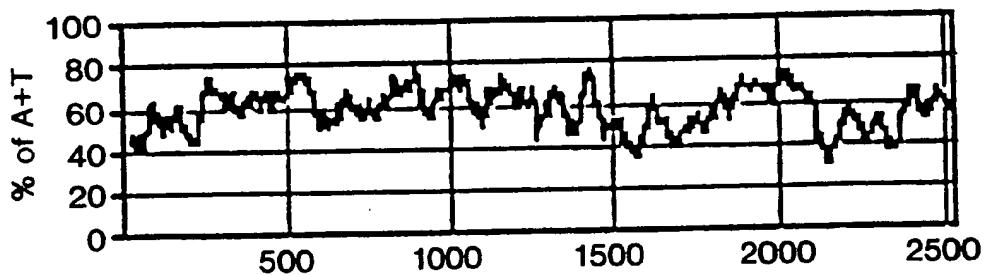


FIG. 22E

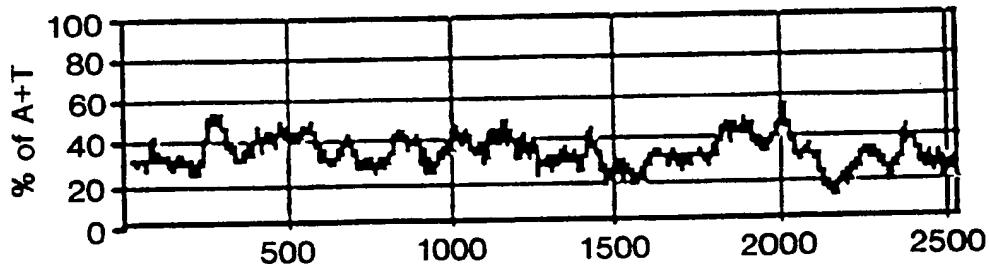


FIG. 22F

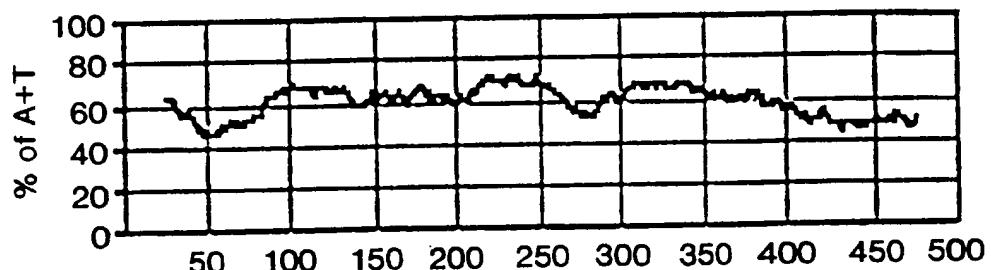


FIG. 22G

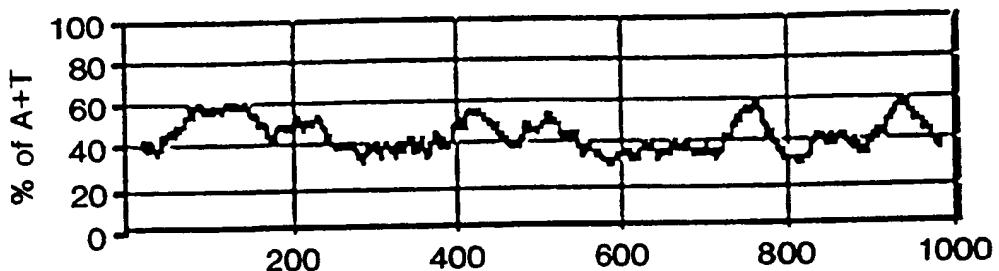


FIG. 22H

gp140.modSF162

gaattcccccaccatggatgcaatgaagagagggtctgtgtgtgtgtggagcagtc
ttcgttcgcccagcgcgtggagaagactgtgggtgaccgttactacggcgtccccgtgtggaaag
gaggccaccaccaccctgttctgcgcacagcacaaggcctacgacaccgagggtgcacaacgtg
tgggccacccacgcctgcgtgcccaccgaccccaaccccaaggagatcgtgtggagaacgtgacc
gagaacttcaacatgtggaaaacaacatggtgagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgacccccctgtgcgtgaccctgcactgcaccaacctg
aagaacgcccaccaacaccaagagcagcaactggaaaggagatggaccgcggcgagatcaagaactgc
agcttcaagggtgaccaccagcatccgcaacaagatgcagaaggagtacgcccgtttctacaagctg
gacgttgtgcccattcgacaacgcacaacaccagctacaagctgtatcaactgcacacaccagcgtgatc
acccaggcctgcccccaaggtgagcttcgagccatccccatccactactgcgc(cccgcggcgttc
gccatcctgaagtgcacgcacaagaagttaacccgcagcggccctgcaccaacgtgagcaccgtg
cagtgcacccacggcatccgcggcgtggtgagcaccagctgtgtgaacggcagcctggccgag
gagggcgtggatccgcagcagaacttaccgcacaacgccaagaccatcatcgtgcagctgaag
gagagcgtggagatcaactgcacccgcggcaacaacacaccgcacagagcatcaccatcgcccc
ggccgcgccttctacgcccaccggcgcacatcatcggtgcacatccgcaggcccactgcacatcagc
ggcgagaagtggaaaacaacaccctgaagcagatcgtgaccaagctgcaggcccagttcgcaacaag
accatcgtgttcaaggcagagcagcggcgcaccccgagatcgtgatgcacagcttcaactgcgc
ggcgagttttctactgcacacagcaccagctgttcaacacgcacccgcacactggaaacaacaccatcgcccc
aacaacaccaacggcaccatcacccgcggcatcaaggcagatcatcaaccgcgtggcaggag
gtgggcacaggccatgtacgcggccatccgcggccagatccgcgtgcagcagcaacatcaccggc
ctgcgtgcacccgcacggcggcaaggagatcagcaacaccaccgcagatcttccgcggccatcgcc
ggcgacatgcgcacaactggcgcagcgagctgtacaagtacaagggtgtgaagatcgagccctg
ggcgtggcccccaccaaggccaagcgcgcgtggcagcgcgagaagcgcgcgtgaccctgggc
gccatgttccctgggttcctggcgcgcggcagcaccatggcgcggcagcgtgaccctgacc
gtgcaggccgcagctgtgcggcatcgtgcagcagcagaacaacctgtgcgcgcacatcgag
gcccagcagcacccgtgcagctgacccgtgtgggcataaggcagatcgtgcaggccgcgtgctggcc
gtggagcgtacctgaaggaccagcagctgtggcatctgggctgcagcggcaagctgatctgc
accaccgcgtgcctggaaacgcgcagctggagcaacaagagcctggaccagatctggaaacaacatg
acctggatggagtggagcgcgcagatcgacaactacaccaacctgtatctacaccctgatcgaggag
agccagaaccagcagcaggagaagaacgcagcaggagctgtggagctggacaagtgccagcctgtgg
aactqgttcgacatcagcaagtggctgtggtacatctaactcgcag

FIG. 23
(SEQ ID NO:36)

gp140.modSF162.delV2

gaattcgccaccatggatgcaatgaagagagggctctgtgtgtgtgtggaggcagtc
ttcgccccccaggcgccgtggagaagactgtgggtgaccgttactacggcgtgcccgtgtggaaag
gaggccaccaccaccctgttctgcgcagcgcacgccaaggcctacgacaccgagggtgcacaacgtg
tgggcacccacgcctgcgtgcccaccgcaccccaaccccaaggagatcgtgtggagaacgtgacc
gagaacttcaacatgtggaaagaacaacatggtgagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccccccgtgcgtgaccctgcactgcaccaacctg
aagaacgccaccaacaccaagagcagcaactgaaaggagatggaccgcggcgagatcaagaactgc
agttcaagggtggcgccggcaagctgtatcaactgcacaccagcgttatcacccaggcctgcccc
aaggtgagcttcgagccatccccatccactactgcgccccccgcgccttcgcattctgtaaagtgc
aacgacaagaagttcaacggcagcggccctgcaccaacgtgagcacccgtgcagtgcacccacggc
atccgccccgtgtgagcacccagctgtgtgaacggcagcctggccgaggaggggcgtggatc
cgcagcggagaacttcaccgacaacgccaagaccatcatcgtgcagctgaaaggagagcgtggagatc
aactgcacccgcggccacaacaacacccgcaagagcatcaccatcgccccccggccgcgccttctac
gccacccggcgacatcatcgccgacatccgcgcaggcccactgcaacatcagcggcgagaagtggaaac
aacaccctgaaggcagatcgtgaccaagactgcaggcccagttcgcaacaagaccatcgtgttcaag
cagagcagcggccggcgaccccgagatcgtgatgcacagcttcaactgcggcggcgagttttctac
tgcaacagcaccctggtcaacagcacctgaaacaacaccatcgcccccaacaacaccaacggc
accatcacccctgcccgcattcaagcagatcatcaaccgcgtggcaggagggtggcaaggccatg
tacgccccccatccgcggccagatccgcgtgcagcagcaacatcaccggcctgtgtgcaccgc
gacggccggcaaggagatcagcaacaccaccgagatcttcgcggccggcgacatgcgcgac
aactggcgcagcggagctgtacaagtacaagggtggtaagatcagccctggcgtggccccacc
aaggccaaggcgccgcgtggcgagcgcgagaagcgcgcggcgtgaccctggcgccatgttccctggc
ttccctggcgccgcggcagaccatggcgccccgcagcctgacccctgaccgtgcaggcccggcag
ctgctgagcggcatcgtgcagcagcagaacaacctgctgcgcgcaccgcggccagcaccctg
ctgcagctgaccgtgtgggcatcaagcagactgcaggcccgcgtgtggcgtggagcgttacctg
aaggaccagcagctgctgggcatctgggctgcagcggcaagctgtatcgcaccaccgcgtgc
tggAACGCCAGCTGGAGCAACAAGAGCCTGGACCAGATCTGGAACAACATGACCTGGATGGAGTGG
GAGCGCGAGATCGACAACACTACACCAACCTGATCTACACCCGTGATCGAGGAGAGGCCAGAACCGAG
GAGAAGAACGAGCAGGAGCTGGAGCTGGACAAGTGGGCCAGCCTGGAACTGGTTCGACATC
AGCAAGTGGCTGTGGTACATCTAACTCGAG

FIG. 24

(SEQ ID NO:37)

gp140.modSF162.delV1V2

gaattcgccaccatggatgcaatgaagagagggtctgtgtgtgtgtgtggaggcagtc
ttcgttcgcccagcgccgtggagaagctgtgggtgaccgttactacggcgtccccgtgaaag
gaggccaccaccaccctgtctgcgcagcgcacgccaaggcctacgacaccgaggtgcacaacgtg
tggccaccacccacgcctgcgtgcccacccgaccccaaccccaaggagatgtgtggagaacgtgacc
gagaacttcaacatgtggagaacaacatggtgagcatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccccccgtgcgtggcgcggcaactgcacgacc
agcgtgatcacccaggcctgccccaaagggtgagcttcgagccatccccatccactactgcgc
gccggcttcgcacccatcgtgaagtgcacacgacaagaagttcaacggcagcggccctgcaccaacgtg
agcaccgtgcagtgcacccacggcatccgccccgtggtgagcacccagctgtgtgaacggcagc
ctggccgaggaggcgtggatccgcagcgagaacttcaccgacaacgccaagaccatcatcgtg
cagctgaaggagagcgtggagatcaactgcacccgccccaaacaacaacacccgcaagagcatcacc
atcggcccccggccgcgcctctacgcacccggcgcacatcatcgccagcatccggcaggccactgc
aacatcagcggcgagaagttgaacaacacccctgaaggcagatgtgaccaagctgcaggcccagttc
ggcaacaagaccatcgtgtcaagcagagcagcggcggcgcaccccgagatcgtgatgcacagcttc
aactgcggcggcgagttctactgcaacagcaccagctgttcaacagcacctggaaacaacacc
atcggccccaacaacacccaaacggcaccatcacoctgcctgcgcataagcagatcatcaaccgc
tggcaggaggtggcaagggcatgtacgccccccatccgcggccagatccgcgcagcagcaac
atcaccggcctgtgtgacccgcacggcggcaaggagatcagcaacaccaccaggatctccgc
cccgccggcggcgacatgcgcacaactggcgcagcgcagctgtacaagtacaagggtggtaagatc
gagccctggcgtggcccccaccaaggccaagcgcgcgtggcagcgcgagaagcgcgcgcgtg
accctggcgcacatgttccctggcttccctggcgcgcggcagcaccatggcgcgcgcagcctg
accctgaccgtgcaggccgcacgtgtgagcggcatcgtgcagcagcagaacaacctgtgcgc
gcacatcgaggcccacgcacgcacccatcgtgcagctgaccgtgtggcatcaagcagatcgc
gtgcgtggccgtggagcgttacctgaaggaccagcagctgtggcatctgggtgcagcggcaag
ctgatctgcaccaccgcgtgcacgcacgtggcaacgcgcacgtggagcaacaagagcctggacc
aacaacatgacccatggatggagtggagcgcgcagatcgcacaactacaccaacctgtatct
atcgaggagagccagaaccaggcaggagaagaacgcgcaggagctgtggagctggacaagtg
acgcctgtggaaactgggtgcacatcgcacgcacgtggctgtggatcatctaactcgag

FIG. 25
(SEQ ID NO:38)

gp140.mut.modsF162

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtctgtgtggaggcagtc
ttcgttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcgtccccgtgtggaaag
gaggccaccaccaccctgttctgcgcacgcacccaaggccatacgacaccgaggtgcacaacgtg
tgggccaccacgcctgcgtgcccacccgaccccaaaaaaaggagatcgtgtggagaacgtgacc
gagaacttcaacatgtggaagaacaacatggtgaggcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccctgtgcgtgaccctgcactgcaccaacctg
aagaacgcccaccaacaccaagagcagcaactggaaggagatggaccgcggcagatcaagaactgc
agcttcaagggtgaccaccagcatccgcaacaagatgcagaaggagtacgcctgttacaagctg
gacgtgggtgcccattcgacaacgcacaccaggatcaagctgatcaactgcacacaccaggcgtgatc
acccaggcctgcccccaaggtgagcttcgagccatccccatccactactgcgcggccgcggcttc
gcacatcctgaagtgcacgcacaagaagttcaacggcagcggccctgcaccaacgtgagcaccgtg
cagtgcacccacggcatccgccccgtggtagcaccaggatcgtgtgaacggcagcctggccgag
gagggcgtggtagccgcagcagaacttcacogacaacgccaagaccatcatcgtgcagctgaag
gagagcgtggagatcaactgcacccgcaccaacacacccgcacagagcatcaccatcgcccc
ggccgcgccttctacgcccacccggcgcacatcatcgccgacatccgcccaggcccactgcacatcagc
ggcgagaagtggaacaacaccctgaaagcagatcgtgaccaagctgcaggcccagttggcaacaag
accatcgtgttcaaggcagagcagcggcggcgcaccccgagatcgtgatgcacagcttcaactgcggc
ggcgaggttttctactgcaacagcacccagctgttcaacagcacctggaacaacaccatcgcccc
aacaacaccaacggcaccatcacccctgcccgcacatcaaggcagatcatcaaccgcgtggcaggag
gtggggcaaggccatgtacgccccccatccgcggccagatccgctgcagcagaacatcaccggc
ctgctgctgacccgcacggcggcaaggagatcagcaacaccaccgagatcttccggccggc
ggcgacatgcgcgacaactggcgcagcgcagctgtacaagggtggtagagatcgagccctg
ggcgtggccccccaccaaggccaagcgcgcgcgtgtgcagcgcgagaagagcgcgcgtgaccctggc
gcacatgttctggcttcctggcgcgcggcagcaccatggcgcggcagcgcgtgaccctgacc
gtgcaggccccgcagctgctgagcggcatcgtgcacgcagaacaacctgctgcgcgcacatcgag
gcccagcagcacctgctgcagctgaccgtgtgggcacatcaaggcagatgcaggcccgcgtgctggcc
gtggagcgcacccatgtacggaccagcagctgtggcattctgggctgcagcggcaagctgatctgc
accaccgcgtgcctggaacgccagctggagcaacaagagcgcctggaccagatctggacaacatg
acctggatggagtgggagcgcgcagatcgacaactacaccaacctgtatctacaccctgatcgaggag
agccagaaccagcaggagaagaacgcgcaggagctgtggagctggacaaagtggccagcctgtgg
aactggttcgacatcagcaagtggctgtggtagatctaaactcgag

FIG. 26

(SEQ ID N0:39)

gp140.mut.modsF162.delV2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgtgtggaggcagtc
ttcgtrtcgcccacgcgcgtggagaagactgtgggtgaccgttactacggcgtgccctgtggaaag
gaggccaccaccaccctgttctgcgcacgcacccaaggccatacgacaccgaggtgcacaacgtg
tggccaccacgcgcgtgcgtgcgcacccacgcacccaaaccccaaggagatcgtgtggagaacgtgacc
gagaacttcaacatgtggaaaacaacatggtgagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccctgtgcgtgaccctgcactgcaccaacctg
aagaacgcaccaacaccaagagcagaactggaaaggagatggaccgcggcagatcaagaactgc
agcttcaagggtggcgccggcaagctgatcaactgcacaccagcgttatcacccaggccctgcccc
aaggtgagcttcgagccatccccatccactactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
aacgacaagaagtcaacggcagcggccctgcaccaacgtgagcaccgtgcagtgcacccacggc
atccgcggcgtggtagcaccagctgtgtgaaacggcagcgcctggcgaggagggcgtggtagc
cgccagcggagaacttcaccgcacaacgcacccatcatcgtgcagctgaaaggagagcgtggagatc
aactgcacccgcaccaacaacaacacccgcacagagcatcaccatcgccgcgcgcgcgcgc
gccaccggcgacatcatcgccgacatccgcggccactgcacacatcagcggcgagaagtggAAC
aacaccctgaagcagatcgtgaccaagctgcaggcccagttcggcaacaagaccatcgtttcaag
cagagcagcggcggcgaccccgagatcgtgatgcacagctcaactgcggcggcgagttttcac
tgcaacacgcacccagctgtcaacagcacctggaaacaacaccatcgcccaacaacaccaacggc
accatcacccctgc
tacgc
gacggccggcaaggagatcagcaacaccaccgcacccatctccgcgcgcgcgcgcgcgc
aactggcgccgc
aaggccaaaggccgcgtggtagcagcgcgagaagagcgcgcgcgcgcgcgcgcgcgc
ttccctggcgccgc
ctgc
ctgc
aaggaccaggcagctgtggcatctgggcgcgcgcgcgcgcgcgcgc
tggaaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
gagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
gagaagaacgcgcgcgcgcgcgcgcgcgcgcgcgc
agcaagtggctgtggatcatctaactcgag

FIG. 27

(SEQ ID NO:40)

gp140.mut.modsF162.delV1V2

gaattcgccaccatggatgcaatgaagagagggtctgtgtgtgtgtgtggagcagtc
ttcgttcgcccagcgccgtggagaagctgtgggtgaccgttactacggcgtgcccgtgtggaaag
gaggccaccaccaccctgttctgcgcacgcacgccaaggctacgacaccgagggtgcacaacgtg
tgggccacccacgcctgcgtgcccacccgaccccaaccccaaggagatcgtgtggagaacagtgcacc
gagaacttcaacatgtggagaacaacatggtgagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccccccgtgcgtggccggcaactgcacgacc
agcgtgatcaccaggcctgcccccaaggtgagcttcgagccatccccatccactactgcgc
gcggcctgcacatcctgaagtgcacgcacaagaagtcaacggcagcggccctgcaccaacgtg
agcacccgtgcagtgcacccacggcatccgcggcgtggtagcaccagctgtgtgaacaggcagc
ctggccgaggaggcgtggtagccgcagcgagaacttcaccgacaacgccaagaccatcatcgtg
cagctgaaggagagcgtggagatcaactgcacccgcggcaacaacaacacccgcaagagcatcacc
atcgccccggccgccttctacgcacccggcagatcatcggcgcacatccgcaggcccactgc
aacatcagcggcggagaagtggaaacaacacccctgaagcagatcgtgaccaagctgcaggcccagttc
ggcaacaagaccatcgtgtcaagcagagcagcggcggcggcggcggcggcggcggcggcggcggcgg
aactcgccgg
atcgcccccaacaacaccaacggcaccatcaccctgcctgcgcacatcaagcagatcatcaaccgc
tggcaggagggtggcaaggccatgtacggccggccggcggcggcggcggcggcggcggcggcggcgg
atcaccggcctgtgtgaccccgacggcggcggcggcggcggcggcggcggcggcggcggcggcgg
cccgccgg
gagcccccgtggcgtggcccccaccaaggccaaaggccacggcggcggcggcggcggcggcggcgg
accctggc
accctgaccgtgcaggcccccggcggcggcggcggcggcggcggcggcggcggcggcggcggcgg
gc
gc
gtgtggccgtggagcgtacctgaaggaccagcagctgtggcgtggcgtggcgtggcgtggcgt
ctgatctgcaccaccggccgtgcctggaaacgcggcggcggcggcggcggcggcggcggcgg
aacaacatgacccgtggatggagtgggagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
atcgaggagagcacaaccaggcaggagaagaacgcgcgcgcgcgcgcgcgcgcgcgc
aqcctgtggaaactgggtgcacatcagcaagtggcgtgtggtacatctaactcgag

FIG. 28

(SEQ ID NO:41)

gp140.mut7.modSF162

gaattcggccaccatggatgcaatgaagagagggctctgctgtgtgctgctgtgtggagcagtc
ttcgttcgcccagcgccgtggagaagctgtgggtgaccgttactacggcgtgccgtgtggaaag
gaggccaccaccaccctgttctgcgcacgcacccaaggcctacgacaccgaggtgcacaacgtg
tggccaccacgcctgcgtgcccaccgaccccaaccccaaggagatcgtgtggagaacgtgacc
gagaacttcaacatgttggagaacaacatggtgagcagatgcacgaggacatcatcagcgtgtgg
gaccagagcctgaagccctgcgtgaagctgacccctgtgcgtgaccctgcactgcaccaacctg
aagaacgcaccaacaccaagagcagaactggaaaggagatggaccgcggcagatcaagaactgc
agcttcaagggtgaccaccaggatccgcaacaagatgcagaaggatgcgcctgttctacaagctg
gacgttgtgcccattcgacaacgacaacaccagctacaagctgatcaactgcaccaacaccagcgtgatc
acccaggcctgccccaaagggtgagcttgcagccatccccatccactactgcgc(cccgcggcttc
gccatcctgaagtgcacgcacaagaagttcaacggcagcggccctgcaccaacgtgagcaccgtg
cagtgcacccacggcatccgccccgtgtgagcaccaggctgtgtgaacggcagcctggccgag
gagggcgtggatccgcagcagaacttaccgcacaacgccaagaccatcatcgtgcagctgaag
gagagcgtggagatcaactgcaccccccccaacaacacaccgcagaagagcatcaccatggcccc
ggccgcgccttctacgccaccggcgacatccgcgcacatccgcgcaggccactgcacatcagc
ggcgagaagtggaaacaacaccctgaagcagatcgtgaccaagctgcaggcccagttcggaacaag
accatcgtgttcaaggcagcagcggcggcgcaccccgagatcgtgatgcacagcttcaactgcggc
ggcgagttttctactgcacacgcaccaggctttcaacacgcaccccttgcaccaacaccatccgc
aacaacaccaacggcaccatcacccctgcgcgcacatcaaggcgcacatccgcgcaggag
gtggcaaggccatgtacgcaccccccattccgcggccagatccgcgcacatccgcgcaggag
ctgcgtgcacccgcacggcggcaaggagatcagcaacacaccaccgcacatccgcgcagg
ggcgacatgcgcacactggcgagcagctgtacaagtacaagggtggtaagatcgagccctg
ggcgtggccccaccaaggccatcagcagcgtggcagagcagaagagcgcgtgaccctggc
gccatgttccctggcttcctggcgcgcggcagcaccatggcgcgcgcagcctgaccctgacc
gtgcaggccccccagctgtgagcggcatcgtgcgcagcagaacaacctgcgcgcacatcgag
gcccagcagcaccctgcgcagctgcgcaggctgtggggcatcaaggcagctgcaggccgcgtgc
gtggagcgcacccctgttggagaccaggcagctgtggggcatctggggctgcagcggcaagctgatctgc
accaccgcgtgcgcacccgtggagcaacaagagcgcgtgaccatctggaaacaacatg
acccggatggagttggagcgcgcagatcgaccaactacaccaacctgtatctacaccctgtatcgaggag
agccagaaccagcaggagaagaacgcaggagctgtggagatggacaagtggccagcctgtgg
aactgggttcgacatcagcaagtggctgtggatcatctaactcgag

FIG. 29
(SEQ ID NO:42)

gp140.mut7.modsF162.delV2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgtgtggagcagtc
ttcgttcgcccacgcgcgtggagaagctgtgggtgaccgttactacggcgtgcccgtgtggaaag
gaggccaccaccaccctgttctgcgcacgcacgccaaggcctacgacacccgagggtgcacaacgtg
tgggccacccacgcctgcgtgcccaccgcaccccaaccccaaggagatcgtgtggagaacgtgacc
gagaacttcaacatgtggagaacaacatgtggagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccccccgtgcgtgaccctgcactgcaccaacctg
aagaacgccaccaacaccaagagcagcaacttggaaaggagatggaccgcggcagatcaagaactgc
agcttcaagggtggcgccggcaagctgatcaactgcacaccaggcgtatcaccaggcctgcccc
aaggtgagcttcgagccatccccatccactactgcgcggcccgccgcttcgcattctgaagtgc
aacgacaagaagttcaacgcgcagcggccctgcaccaacgtgagcaccgtgcagtgcacccacggc
atccgcggccgtggtagcaccaggctgtgtgaacggcagcctggccgaggaggcgtggtagc
cgcagcggagaacttcaaccgacaacgcaccaagaccatcatcgtgcagctgaaggagagcgtggagatc
aactgcacccggcccaacaacaacacccgcacagagcatcaccatcgcccccggccgcgccttctac
gccaccggcgacatcatcgccgacatccgcaggcccactgcacatcagcggcgagaagtggaaac
aacaccctgaagcagatcgtgaccaagctgcaggcccagttcgccaaacagaccatcgtgttcaag
cagagcagcggccggcgaccccgagatcgtgatgcacagcttcaactgcggccggcgagttcttctac
tgcaacagcaccaggctgttcaacagcacctggacacaacaccatcgcccccggccgcgccttctac
accatcaccctgcgcgcacatcaaggctggcaggagggtggcaaggccatg
tacgccccccatccgcggccagatccgcgtgcagcagcaacatcaccggcctgtgtgaccggc
gacggccggcaaggagatcagcaacaccaccgagaatctccgcggccggccgcacatgcgcac
aactggcgcagcgtgtacaaggctgtgtacaaggatcagcggccatgttccctggcc
aaggccatcagcagcgtggtagcagagcggagaagagcgcgcgtgaccctggccgcacatgttccctggcc
ttccctggccgcggccggcagcaccatggccggccgcagccgaccctgaccgtgcaggccggcc
ctgctgagcggcatcgtgcagcagcagaacaaccgcgtgcgcgcacatcgaggcccagcagcac
ctgcagctgaccgtgtggccatcaagcagctgcaggcccgcgtgctggccgtggagcgcgtac
aaggaccaggcagctgtggcatctggccgtgcagcggcaagctgtgtgcaccaccgcgtgccc
tggAACGCCAGCTGGAGCAACAAGAGCCTGGACCAGATCTGGAACACATGACCTGGATGGAGTGG
GAGCGCGAGATCGACAACACACCCGTGATCTACACCCGTGACCGAGCTGGAGCAGAACCAGCAG
GAGAAGAACGAGCAGGAGCTGGAGCTGGAGCAGTGGGCCAGCCTGTGGAACCTGGTTCGACATC
AGCAAGTGGCTGTGGTACATCTAACTCGAG

FIG. 30
(SEQ ID NO:43)

gp140.mut7.modsF162.delV1V2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgctgctgctgtggagcagtc
ttcgttcgcccagcgccgtggagaagctgtgggtgaccgttactacggcgtgcccgtgtggaaag
gaggccaccaccaccctgtctgcgcagcgcacgccaaggcctacgacaccgaggtgcacaacgtg
tggccaccaccacgcctgcgtgcccaccgaccggccaaaaacccccaggagatcgctgtggagaacgtgacc
gagaacttcaacatgtggaaaacaacatggtgagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccccccgtgcgtgggcgcggcaactgcgcagacc
agcgttatcaccaggcctgccccaaaggtagcttcgagccatccccatccactactgcgcgggg
gccggcttcgcacatcctgaagtgcacgcacaagaagtcaacggcagcggccctgcaccaacgtg
agcacccgtgcagtgcacccacggcatccgcggccgtggtagaccccgactgctgtgaacggcagc
ctggccgaggaggcgtggtagccgcagcggaaacttcaccgacaacgccaagaccatcatcgtg
cagctgaaggagagcgtggagatcaactgcacccggcccaacaacaacacccgcaagagcatcacc
atcggcccccggccgcgccttcacgcacccggcagacatcatcgccgacatccggcaggccactgc
aacatcagcggcgagaagtggaaacaacaccctgaaggcagatcgtgaccaagctgcaggcccgatcc
ggcaacaagaccatcgtgtcaagcggcagcggccggcggccggcggccggcggccggcggccggc
aactgcggcggcgagttcttactgcaacagcaccaggctgttcaacagcacttggaaacaacacc
atcggccccaacaacaccaacggcaccatcaccctgcgcgcacatcaaggcagatcatcaaccgc
tggcaggaggtggcaaggccatgtacggcccccacatccgcggccagatccgcgtcagcggc
atcaccggcctgtgtgacccggcggccggcggccggcggccggcggccggcggccggcggccggc
ccggccggccggcggcggccggcggccggcggccggcggccggcggccggcggccggcggccggc
ggccggccggccggcggccggcggccggcggccggcggccggcggccggcggccggcggccggc
accctggcggccatgttccctggcttccctggcggccggccggcggccggcggccggcggccggc
accctgaccgtgcaggcccgccagctgtgagcggcatcgtcagcggcagaacaacactgtgcgc
gccatcgaggcccagcagcggccacatcgatcgactgtggccatcaaggcagatcgatcgaggcccgc
gtgctggccgtggagcgtacatgtggccatcgatcgatcgatcgatcgatcgatcgatcgatcg
atcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
aacaacatgacactgtggatggagtgggagcgcgcagatcgacactacaccacccgttatc
atcgaggagagccagaaccaggcaggagaagaacaggcaggagcaggagctgtggagctggacaat
agcctgtggaaactgtgttcgcacatcgacatcgacatcgacatcgacatcgacatcgacatcg

FIG. 31

(SEQ ID NO:44)

gp140.mut8.modSF162

gaattggccaccatggatgcaatgaagagagggtctgtgtgtgtgtgtggagcagtc
ttcgccccagcgcgtggagaagctgtgggtgaccgttactacggcgtcccgtgtggaaag
gaggccaccaccaccctgttctgcgcagcgcacgccaaggctacgacaccgaggtgcacaacgtg
tggccaccacgcctgcgtgcccaccgaccccaaccccaaggagatcgtgtggagaacagtgcacc
gagaacttcaacatgtggagaacaacatggtgagcagatgcacgaggacatcatcagcgtgtgg
gaccagagcctgaagccctgcgtgaagctgaccctgtgcgtgaccctgcactgcaccaacctg
aagaacgcaccaacaccaagagcagcaactggaaaggagatggaccgcggagatcaagaactgc
agcttcaagggtgaccaccagcatccgcaacaagatgcagaaggatcgcctgttctacaagctg
gacgtggccatcgacaacgcacaacaccagctacaagctgtatcaactgcacaccagcgtgatc
acccaggcctgccccaaagggtgagcttgcagccatccccatccactactgcgcggccgcggcttgc
gcacatcctgaagtgcacgcacaagaagttcaacggcagcggccctgcaccaacgtgagcaccgtg
cagtgcacccacggcatccgccccgtgtgagcaccagctgtgtgcacggcagcgtggccag
gagggcgtggatccgcagcagaacttcaccgcacaacgccaagaccatcatcgtgcagctgaag
gagagcgtggagatcaactgcacccggccaaacaacaacacccgcacagagcatcaccatcgcccc
ggccgcgccttctacgccacggcgacatcatcggtgcacatccgcccactgcacatcagc
ggcgagaagtggaaacaacaccctgaagcagatcgtgaccaagctgcaggcccagttcgcaacaag
accatcgtgttcaagcagagcagcggccgcaccccgagatcgtgtatgcacagcttcaactgcgc
ggcgagttttctactgcacacagcaccagctgttcaacagcacctggaaacaacaccatcgcccc
aacaacaccaacggcaccatcaccctgcgcacatcaagcagatcatcaaccgctggcaggag
gtgggcacaggccatgtacggcccccattccggccgcacatccgcgtgcagcagacatcaccggc
ctgcgtgtgcacccgcacggcggcaaggagatcagcaacaccaccgagatcttccggccggc
ggcgacatgcgcgacaactggcgacgcgcagctgtacaagtacaagggtgtgaagatcgagccctg
ggcgtggcccccaccatcgccatcagcagcgtgtgcagagcggagaagagcgcgtgaccctggc
gcacatgttctggcttccctggcgccggcagcaccatggcgccgcagcctgcacccctgacc
gtgcaggccgcacgcgtgtgcagcggcatcgtgcagcagcagaacaacctgtgcgcgcacatcgag
gcccagcagcacctgtgcagctgaccgtgtgggcacatcaagcagctgcaggcccgcgtgtggcc
gtggagcgtacccatggacaggaccagcagctgtggcatctgggtgcagcggcaagctgtatctgc
accaccggcgtgcctggacggccacgtggagcacaagagcctgggaccagatctggacaaacatg
acctggatggagtggagcgcgcagatcgacaactacccaacctgtatctacaccctgtatcgaggag
agccagaaccagcaggagaagaacgcagcaggagctgtggagctggacaaagtggccagcgtgtgg
aactqqtctgacatcagcaagtggctgtggtacatctaactcgcag

FIG. 32
(SEQ ID NO:45)

44 / 131

gp140.mut8.modSF162.delV2

gaattcgccaccatggatgcaatgaagagagggtctgtgtgtgtgtgtgtggagcagtc
ttcgcccccccgccgtggagaagctgtgggtgaccgtgtactacggcgtgcccgtgtggaaag
gaggccaccaccaccctgttctgcgccagcgacgccaaggcctacgacaccgaggtgcacaacgtg
tggccaccaccacgcctgcgtgcccaccgaccccaaccccccaggagatcgtgtggagaacgtgacc
gagaacctcaacatgtggaaagaacaacatggtgagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccctgtgcgtgaccctgcactgcaccaacctg
aagaacgcaccaacaccaagagcagcaactggaaaggagatggaccgcggcgagatcaagaactgc
agcttcaaggtggcgccggcaagctgtatcaactgcaacaccagcgtatcaccaggcctgcccc
aaggtgagcttcgagccatccccatccactactgcgcccccgccgttcgcacatcctgaagtgc
aacgacaagaagtcaacggcagcgccccctgcaccaacgtgagcacgcgtgcagtgcacccacggc
atccgcggccgtggtagcaccctgcgtgtggaaacggcagcctggccgaggaggcggtgtgatc
cgcagcagaacttcaccgacaacgcacaagaccatcatcgtgcagctgaaggagagcgtggagatc
aactgcacccgcoccaacaacaacacccgcaagagcatcaccatcgcccccgccgcgcctctac
gccacccggcgacatcatcgccgacatccgcaggcccactgcacatcagcgccgagaagtggAAC
aacaccctgaagcagatcgtgaccaagctgcagccccagtcggcaacaagaccatcgtgttcaag
cagagcagcggcgccgaccccgagatcgtgatgcacagctcaactgcggcgccgagttttctac
tgcaacacgcacccagctgttaacacgcacccgtggaaacaacaccatcgcccccaacaacaccaacggc
accatcaccctgcctgcgcatcaagcagatcatcaaccgcgtggcaggagggtggcaaggccatg
tacgcggccatcccgccggccagatccgcgtgcagcagcaacatcaccggcctgcgtgcacccgc
gacggccggcaaggagatcagcaacaccaccgagatcttccgcggccggccgacatgcgcac
aactggcgcagcggagctgtacaagtacaagggtggtagagatcgagccctggcggtggcccccacc
atcgccatcagcagcgtggtagcagagcggagaagagcgcgtgaccctggcgccatgttccctggc
ttcctggcgccggccggcagcaccatggcgcccgagccgaccctgaccgtgcaggccccagcagcac
ctgcgtgagcggcatcgtgcagcagcagaacaacctgcgtgcgcgcacatcgaggcccgac
ctgcagctgaccgtgtgggcatcaagcagctgcaggccccgtgcgtggcggtggagcgcgtac
aaggaccaggcagctgtggcatctggggctgcagcggcaagctgtatcgcaccaccgcgtgccc
tggaaacgcagctggagcaacaagagccgtggaccagatctggaaacaacatgacactggatggagtg
gagcgcgagatcagacaactacaccaacctgatctacaccctgatcgaggagagccagaaccagcag
gagaagaacgagcaggagctgtggagctggacaagtggccagcctgtggactggatcgacatc
agcaagtggctgtggatcatctaactcgag

FIG. 33

(SEQ ID NO:46)

gp140.mut8.modsF162.delV1V2

gaattcgccaccatggatgcaatgaagagagggctctgtgtgctgtgtgtggaggcagtc
ttcgccccagcgccgtggagaagctgtgggtgaccgtgtactacggcgtgcccgtgtggaaag
gaggccaccaccaccctgttctgcgccagcgacgccaaggcctacgacaccgaggtgcacaacgtg
tggccaccacccacgcctgcgtgcccaccgaccccaaccccccaggagatcgtgtggagaacgtgacc
gagaacttcaacatgtggaaagaacaacatggtggagcagatgcacgaggacatcatcagcctgtgg
gaccagagcctgaagccctgcgtgaagctgaccccccgtgcgtggcgccggcaactgccagacc
agcgtgatcacccaggcctgcccccaaggtgagctcgagccatccccatccactactgcgcccccc
gcccggctcgccatcctgaagtgcacgacaagaagttcaacggcagcgccccctgcaccaacgtg
agcaccgtgcagtgcacccacggcatccgccccgtggtgagcaccaggctgtgtgaacggcagc
ctggccgaggaggcggtggatccgcagcgagaacttaccgacaacgccaagaccatcatcgtg
cagctgaaggagagcgtggagatcaactgcaccccccacaaacaacacccgcacagagcatcacc
atcggcccccggccgccttctacgccaccggcgacatcatcgccgacatccgcccaggccactgc
aacatcagcgccgagaagtgaaacaacaccctgaagcagatcgtgaccaagctgcaggcccagttc
ggcaacaagaccatcgtgttcaagcagagcagcgccggcgaccccgagatcgtgatgcacagcttc
aactgcggcgccgagttcttactgcacagcaccaggctgttcaacagcaccttggaaacaacacc
atcggccccaacaacaccaacggcaccatcaccctgcccgcacatcaagcagatcatcaaccgc
tggcaggagggtggcaaggccatgtacgccccccatccgcggccagatccgtgcagcagcaac
atcaccggcctgctgctgacccgcacggcgcaaggagatcagcaacaccaccaggatctccgc
cccgccggcgccgacatgcgcgacaactggcgagcgtgtacaagtacaagggtggtaagatc
gagccctggcgccatgttcttggcttctggcgccggcagcaccatggcgccggcagcctg
accctgaccgtgcaggcccggcagctgtgagcggcatcgtgcagcagcagaacaacctgtgcgc
ccatcgaggcccagcagcacctgtgcagctgaccgtgtgggcatcaagcagctgcaggcccgc
gtgctggccgtggagcgctacctgaaggaccagcagctgtggcatctgggctgcagcggcaag
ctgatctgcaccaccgcgtgcccgtggaaacgccaagctggagcaacaagagcctggaccagatctgg
aacaacatgacctggatggagtgggagcgcgagatcgacaactacaccaacctgtatctacaccctg
atcgaggagagccagaaccagcaggagaagaacgagcaggagctgtggagctggacaagtgggca
agcctgtggactggatcgacatcagcaagtggctgtggatcatctaactcgag

FIG. 34

(SEQ ID NO:47)

gp160.modSF162

gaattcgccaccatggatgcaatgaagagagggctgtgtgtgtgtgtggaggcagtc
 ttcgtttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcgtgcccgtgtggaaag
 gaggccaccaccaccctgttctgcgccagcgcacgccaaggcctacgcacaccgaggtgcacaacgtg
 tgggccacccacgcctgcgtgcccaccgaccccaaccccccaggagatcgtgtggagaacgtgacc
 gagaacttcaacatgtggaaagaacaacatggtgagcagatgcacgaggacatcatcagcctgtgg
 gaccagagcctgaagccctgcgtgaagctgacccctgtgcgtgaccctgcactgcaccaacctg
 aagaacgccaccaacaccaagagcagcaacttggaaaggagatggaccgcggcgagatcaagaactgc
 agcttcaaggtgaccaccagcatccgcacaacaagatgcagaaggagtacgcctgttacaagctg
 gacgtggtgcctcatgcacaacgcacaacaccagctacaagctgtatcaactgcacaccagcgtgatc
 acccaggcctgccccaggtagttagcttgcgcggccatccccatccactactgcgcggccggcttc
 gccatcctgaagtgcacgcacaagaagttaacggcagcggccctgcaccaacgtgagcaccgtg
 cagtgcaccccacggcatccgcggctggtgagcaccaggctgtgtgacggcactgcgcggccggag
 gagggcgtggatccgcagcgagaacttacccgacaacgcaccaagaccatcatcgtgcagctgaag
 gagagcgtggagatcaactgcacccggccaaacaacaacacccgcaagagcatcaccatggggccc
 ggccgcgccttctacgcacccgcacatcatcgccgacatccgcaggcccactgcacatcagc
 ggcgagaagtggaaacaacaccctgaagcagatcgtgaccaagctgcaggcccagttcggcaacaag
 accatcgttcaagcagacgcggcggcgcacccgagatcgtgatgcacagcttcaactgcggc
 ggcgagtttttactgcaacacgcacccagcttcaacacgcacccatggggccc
 aacaacaccaacgcacccatcaccctgcctgcgcacatcaagcagatcatcaaccgcgtggcaggag
 gtgggcaggccatgtacgc
 ctgc
 ggcgcacatgcgcgacactggcgcagcgagctgtacaagtcacagggtgtgaagatcgcaggccc
 ggcgtggccccccaccaaggccaagcgcgcgtggcagcgcgagaagcgcgcgcgcgcgcgcgc
 gccatgttctggcttcctggc
 gtgcaggccccccagctgc
 gcccagcgcacccgc
 gtggagc
 accaccgc
 acctggatggagttggc
 agccagaaccaggcaggagaagaacgcgcgcgcgcgcgcgcgcgcgcgcgcgc
 aactgggtcgacatcagcaagtggctgtgttatcaagatcttcatcatgatcgtggcggccctg
 gtgggc
 ctgagcttcagacccgcgttccccgcgcgcgcgcgcgcgcgcgcgcgcgc
 ggccggc
 gacctgc
 atcgtggagctgtggccgcgcgcgcgcgcgcgcgcgcgcgc
 tggatcaggagctgaagaacacgcgcgcgcgcgcgcgc
 ggcaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
 atccgcgcaggcgcgcgcgcgcgcgcgcgcgcgcgcgc

FIG. 35
(SEQ ID NO:48)

gp160.modSF162.delV2

gaattcgccaccatggatgcaatgaagagagggctctgtgtgtgtgtgtggaggcagtc
 ttcgtttcgcccacgcgcgtggagaagactgtgggtgaccgtgtactacggcgtgcccgtgtggaaag
 gaggccaccaccacccctgttctgcgcacgcacccaaggccatacgacaccgagggtgcacaacgtg
 tgggccaccacacgcctgcgtgcccaccgaccccaaccccccaggagatcgtgtggagaacgtgacc
 gagaacctcaacatgtggaaagaacaacatggtgagcagatgcacgaggacatcatcagcctgtgg
 gaccagagcctgaagccctgcgtgaagctgaccctgtgcgtgaccctgcactgcaccaacctg
 aagaacgcaccaacaccaagagcagcaactggaaaggagatggaccgcggcagatcaagaactgc
 agcttcaagggtggcgccggcaagctgatcaactgcaacaccagcgtgatcacccaggcctgcccc
 aaggtgagcttcgagccatccccatccactactgcgc(cccgcggcgcgcgcgcgcgcgcgcgc
 aacgacaagaagtcaacggcagcggccctgcaccaacgtgagcaccgtgcagtgcacccacggc
 atccgc(cccgtggtgagcaccctgactgtgtgaacggcagcctggcgaggaggcgtgtgatc
 cgcagcggagaacttcaccgacaacgccaagaccatcatcgtgcagctgaaggagagcgtggagatc
 aactgcacccgc(cccaacaacacacccgcaagagcatcaccatcgcccccggccgcgccttctac
 gccacccggcgacatcatcgccgacatccgcaggcccactgcaacatcagcggcggagaagtggAAC
 aacaccctgaagcagatcgtgaccaagactgcagggccagttcgccacaagaccatcgtgttcaag
 cagagcagcggccggcgcaccccgagatcgtgatgcacagctcaactgcggcggcggagttcttctac
 tgcaacacgcacccagctgttcaacacgcacccctggaaacaacaccatcgcccccaccaacaccaacggc
 accatcacccctgc(cctgcgcacatcaaggcagatcatcaaccgcgtggcaggagggtggcaaggccatg
 tacgc(ccccatccgcggccagatccgcgtgcagcaacatcaccggcctgctgtgaccccg
 gacggcggcaaggagatcagcaacaccaccgagatctccgc(cccggccggcgcacatgcgcgc
 aactggcgcagcggagctgtacaagtacaagggtgtgaagatcgagccctggcgtggcccccacc
 aaggccaagcgccgcgtggcagcgcgagaagcgcgcgtgaccctggcgcgcgcgcgcgcgcgcgc
 ttccctggcgccgc(ccggcagcaccatggcgcccgacgcctgaccctgaccgtgcaggcccgccag
 ctgctgagcggcatcgtgcagcagcagaacaactctgctqgcgcgcgcgcgcgcgcgc
 ctgcagctgaccgtgtgggcatcaaggcagactgcgcaggcccgcgtgctggccgtggagcgcgc
 aaggaccaggcagctgtggcatctgggctgcagcggcaagctgatcgcaccaccgcgcgc
 tggaaacgcagctggagcaacaagagaccctggaccagatctggaaacaacatgacccctggatggagtg
 gagcgcgagatcagacaactacaccaacctgatctacaccctgatcgaggagagccagaaccaggc
 gagaagaacgcagcggaggactgtggagactgtggacaagactggccagcctgtggaaactgggtcgacatc
 agcaagtggctgtggatcatcaagatcttcatcatgatcgtggcgccctggcggccctggcgc
 gtgttcaaccgtgtgagcatcgtgaaccgcgtgcgcaccggctacagcccccctgagcttccagacc
 cgcttccccgc(cccccgccggccgcgcgcgcgcgcgcgcgcgcgcgc
 cgcgcaccgcagcagccctggcgcacggcctgcgcgcgcgcgcgcgcgcgc
 tgccctgttcaaccaccgcctgcgcgcgcgcgcgcgcgcgcgcgc
 ggccgc
 aagaacacgcgcgcgtgagcctgttgcacgcacatgcgcgcgcgcgc
 atcgaggtggcccccagcgcacatcgccgcgcgcgcgcgcgcgc
 gagc

FIG. 36
(SEQ ID NO:49)

48 / 131

gp160.modSF162.delV1V2

FIG. 37
(SEQ ID NO:50)

gp120wtUS4

ACAAACAGTCTTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAAGAAG
CAACCACCACTCTGTTTGTGCATCAGATGCTAAAGCATACAAAGCAGAGGC
ACATAACGTCTGGGCTACACATGCCTGTGTACCCACAGACCCCCAACCCACAG
GAAGTAAATTAAACAAATGTGACAGAAAATTAAACATGTGGAAAAATAACA
TGGTGGAACAGATGCATGAGGATATACTAGTTATGGGATCAAAGCCTAAA
GCCATGTGAAAATTAAACCCACTCTGTGTTACTTTAAATTGTACTGATAAGT
TGACAGGTAGTACTAATGGCACAAATAGTACTAGTGGCACTAATAGTACTAG
TGGCACTAATAGTACTAGTACTAATAGTACTGATAGTTGGGAAAAGATGCCA
GAAGGAGAAATAAAAAGTCTTCAATATCACCACAAGTGTAAAGAGATA
AAGTGCAGAAAGAATATTCTCTCTTCTATAAAACTGATGTAGTACCAATAGAT
AATGATAATGCTAGCTATAGATTGATAATTGTAATACCTCAGTCATTACACA
AGCCTGTCCAAGGTATCTTGAACCAATTCCCACATATTGTGCCCGG
CTGGTTTGCATTCTAAAGTGTAAAGATAAGAAGTTCAATGGAACAGGACC
ATGTAAAAATGTCAGCACAGTACAATGCACACATGGAATTAGACCAGTAGTA
TCAACTCACTGCTGTTAAATGGCAGTCTAGCAGAAGAAGAGATAGTACTTA
GATCTGAAAATTTCACAGACAATGCTAAAACCATATAATAGTACAGCTGAATGA
ATCTGTAGAAATTAAATTGTATAAGACCCAAACAATAATACAAGAAAAAGTATA
CATATAGGACCAGGGAGAGCATTATGCAACAGGTGATATAATAGGAGACA
TAAGACAAGCACATTGTAACATTAGTAAAGCAAATGGACTAACACTTACA
ACAGATAGTTGAAAAATTAAAGAGAACAAATTGGGAAATAATAAAACAATAATC
TTAATTCACTCCTCAGGGAGGGGACCCAGAAATTGTATTCACTTAAATTG
TGGAGGGGAAATTCTATTGTAATACATCACAATTTAAATTAGTACCTGGA
ATATTACTGAAGAGGTAAATAAGACTAAAGAAAATGACACTATCATACTCCC
ATGCAGAATAAGACAAATTATAACATGTGGCAAGAAGTAGGAAAAGCAAT
GTATGCCCTCCCATCAGAGGACAAATTAAATGTTCATCAAATATTACAGGG
CTGCTATTAACTAGAGATGGTGGTACTAACAAATAATAGGACGAACGACACCG
AGACCTTCAGACCTGGGGAGGAAACATGAAGGACAATTGGAGAAGTGAAT
TATATAAATATAAAGTAGTAAGAATTGAACCATTAGGAGTAGCACCCACCC
GGCAAAGAGAACAGTAGGTGCAAAGAGAGAAAAGA

FIG. 38

(SEQ ID NO:51)

gp140wtUS4

ACAACAGCTTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAAGAAG
 CAACCACCACTCTGTTTGTGCATCAGATGCTAACAGCATACAAAGCAGAGGC
 ACATAACGTCTGGGCTACACATGCCCTGTGTACCCACAGACCCCACCCACAG
 GAAGTAAATTAAACAAATGTGACAGAAAAATTAAACATGTGGAAAAATAACA
 TGGTGGAACAGATGCATGAGGATATAATCAGTTATGGGATCAAAGCCTAAA
 GCCATGTGTAAAATTAAACCCACTCTGTGTACTTTAAATTGTACTGATAAGT
 TGACAGGTAGTACTAATGGCACAAATAGTACTAGTGGACTAATAGTACTAG
 TGGCACTAATAGTACTAGTACTAATAGTACTGATAGTTGGGAAAAGATGCCA
 GAAGGAGAAATAAAAACTGCTCTTCAATATCACCACAAGTGTAAAGAGATA
 AAGTGCAGAAAGAATATTCTCTTCTATAAAACTGATGTAGTACCAATAGAT
 AATGATAATGCTAGCTATAGATTGATAATTGTAATACCTCAGTCATTACACA
 AGCCTGTCCAAAGGTATCTTTGAACCAATTCCACATCATTATTGTGCCCGG
 CTGGTTTGCATTCTAAAGTGTAAAGATAAGAAGTTCAATGGAACAGGACC
 ATGTAAAAATGTCAGCACAGTACAATGCACACATGGAATTAGACCAAGTAGTA
 TCAACTCAACTGCTGTAAATGGCAGTCTAGCAGAAGAAGAGATAGTACTTA
 GATCTGAAAATTTCACAGACAATGCTAAAACCATAATAGTACAGCTGAATGA
 ATCTGTAGAAATTAAATTGTATAAGACCCAACAATAATACAAGAAAAAGTATA
 CATATAGGACCAGGGAGAGCATTATGCAACAGGTGATATAATTAGGAGACA
 TAAGACAAGCACATTGTAACATTAGTAAAGCAAACGGACTAACACTTTAGA
 ACAGATAGTTGAAAATTAAAGAGAACAAATTGGAATAATAAAACAATAATC
 TTTAATTCCATCCTCAGGAGGGACCCAGAAATTGTATTTCACAGTTAATTG
 TGGAGGGAAATTCTATTGTAATACATCACAACATTAAATAGTACCTGGA
 ATATTACTGAAGAGGTAAATAAGACTAAAGAAAATGACACTATCATACTCCC
 ATGCAAGATAAGACAAATTAAACATGTGGCAAGAAGTAGGAAAAGCAAT
 GTATGCCCTCCCATCAGAGGACAAATTAAATGTCATCAAATATTACAGGG
 CTGCTATTAACTAGAGATGGTGGTACTAACAAATAAGGACGAACGACACCG
 AGACCTTCAGACCTGGGGAGGAAACATGAAGGACAATTGGAGAAGTGAAT
 TATATAAATATAAGTAGTAAGAATTGAAACCAATTAGGAGTAGCACCCACCC
 GGCAAAGAGAACAGTGGTGCAAAGAGAGAAAAGAGCAGTGGACTAGGAG
 CTTTGTTCATTGGGTTCTGGGAGCAGCAGGAAGCACTATGGCGCAGCGTC
 AGTACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAAGTCAACAG
 CAGAACAAATTGCTGAGAGCTATTGAGGCGAACAGCATCTGTCACACTCA
 CGGCTGGGCATCAAACAGCTCCAGGCAAGAACCTGGCTGTGGAAAGATA
 CCTAAAGGATCAACAGCTCTAGGGATTGGGTTGCTCTGGAAAACCTATT
 GCACCAACTACTGTGCCTGGAACTCTAGTTGGAGTAATAATCTGACTGAG
 ATTGGATAATATGACCTGGATGGAGTGGAAAAGAGAAAATTGGCAATTATA
 CAGGCTTAATATAACATTAAATTGAAATAGCACAAACAGCAAGAAAAGAA
 TGAACAAGAATTATTGAAATTAGACAAAGTGGCAAGTTGTGGATTGGTT
 GATATAACAAACTGGCTGTGGTATATA

FIG. 39
(SEQ ID NO:52)

gp160wtUS4

ACAACAGTCTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAAGAAG
CAACCACCACTCTGTTTGTGCATCAGATGCTAAAGCATACAAAGCAGAGGC
ACATAACGTCTGGCTACACATGCCTGTGTACCCACAGACCCCCAACCCACAG
GAAGTAAATTAAACAAATGTGACAGAAAATTAAACATGTGGAAAAATAACA
TGGTGGAACAGATGCATGAGGATATAATCAGTTATGGGATCAAAGCCTAAA
GCCATGTGTAAAATTAAACCCCACCTGTGTTACTTTAAATTGTACTGATAAGT
TGACAGGTAGTACTAATGGCACAAATAGTACTAGTGGCACTAATAGTACTAG
TGGCACTAATAGTACTAGTACTAATAGTACTGATAGTTGGAAAAGATGCCA
GAAGGAGAAATAAAACTGCTCTTCAATATCACCACAAGTGTAAAGAGATA
AAGTGCAGAAAGAATATTCTCTCTATAAACTTGATGTAGTACCAATAGAT
AATGATAATGCTAGCTATAGATTGATAACCTCAGTCATTACACA
AGCCTGTCCAAGGTATCTTGAACCAATTCCCATACATTATTGTGCCCGG
CTGGTTTGCATTCTAAAGTGTAAAGATAAGAAAGTTCAATGGAACAGGACC
ATGTAAAAATGTCAGCACAGTACAATGCACACATGGAATTAGACCAAGTAGTA
TCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAAGAAGAGATAGTACTTA
GATCTGAAAATTTCACAGACAATGCTAAAACCATAATAGTACAGCTGAATGA
ATCTGTAGAAATTAAATTGTATAAGACCCACAATAATACAAGAAAAAGTATA
CATATAGGACCAGGGAGAGCATTTATGCAACAGGTGATATAATAGGAGACA
TAAGACAAGCACATTGTAACATTAGTAAAGCAAACATTGGGATAATAAAACAATAATC
TTAATTATCCTCAGGAGGGACCCAGAAATTGTATTTCACAGTTAATTG
TGGAGGGAAATTCTATTGTAAATACATCACAACATTAAATAGTACCTGG
ATATTACTGAAGAGGTAATAAGACTAAAGAAAATGACACTATCATACTCCC
ATGCAAGATAAGACAAATTATAACATGTGGCAAGAAGTAGGAAAAGCAAT
GTATGCCCTCCCATCAGAGGACAAATTAAATGTTCATCAAATATTACAGGG
CTGCTATTAACTAGAGATGGTGGTACTAACAAATAATAGGACGAACGACACCG
AGACCTTCAGACCTGGGGAGGAAACATGAAGGACAATTGGAGAAGTGAAT
TATATAAAATAAAAGTAGTAAGAATTGAACCATTAGGAGTAGCACCCACCC
GGCAAAGAGAAGAGTGGTGCAGAGAGAAAAGAGCAGTGGACTAGGAG
CTTGTTCATTGGTTCTTGGGAGCAGCAGGAAGCAGTATGGCGCAGCGTC
AGTACGCTGACGGTACAGGCCAGACAATTATGTCTGGTATAGTGCACACAG
CAGAACAAATTGCTGAGAGCTATTGAGGCGCAACAGCATCTGTTGCAACTCA
CGGTCTGGGCATCAAACAGCTCCAGGCAAGAACCTGGCTGTGGAAAGATA
CCTAAAGGATCAACAGCTCTAGGGATTGGGTTGCTCTGGAAAACCTATT
GCACCACTACTGTGCCTTGGAACTCTAGTGGAGTAATAATCTGACTGAG
ATTGGGATAATATGACCTGGATGGAGTGGAAAAGAGAAAATTGGCAATTATA
CAGGCTTAATATAACAATTAAATTGAAATAGCACAAACCCAGCAAGAAAAGAA
TGAACAAAGAATTATTGGAATTAGACAAGTGGCAAGTTGTGGAATTGGTT
GATATAACAAACTGGCTGTGGTATATAAGAATATTCTATAATGATAGTAGGAG
GCTTGATAGGTTAAGAATAGTTTGCTGTACTTTCTATAAGTGAATAGAGTT
AGGCAGGGACTCACCAATATCATTGCAGACCCGCCTCCAGCTCAGAGGG

FIG. 40A
(SEQ ID NO:53)

GACCCGACAGGCCGAAGGAATCGAAGAAGAAGGTGGAGAGAGAGACAGA
GACAGATCCAATCGATTAGTGCATGGATTATTGGCACTCATCTGGGACGATCT
GCGGAGCCTGTGCCTCTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTG
TAGCGAGGATTGTGGAACCTCTGGGACGCAGGGGTGGAAGCCCTCAAGTA
TTGGTGGAAATCTCCTGCAGTATTGGAGTCAGGAGCTAAAGAGTAGTGCTGTT
AGTTTGTAAATGCCACAGCAATAGCAGTAGCTGAAGGGACAGATAGGATTA
TAGAAATAGTACAAGAATTTTAGAGCTGTAATTCACATACCTAGAAGAAT
AAGACAGGGCTGGAGAGAGGGCTTACTATAA

FIG. 40B
(SEQ ID NO:53)

gp120.modUS4

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
GTCTCGTTGCCCAAGGCCACCACCGTGTGGGTGACCGTGTACTACGGCGTCCCCGTG
TGGAGGAGGCCACCACCCACCGTGTGGCCAGCGACGCCAAGGTTACAAGGCCGAGGC
CCACAAACGTGTGGGCCACCCACGCCCTGCGTGTGCCCACCGACCCCAACCCCCAGGAGGTGAACC
TGACCAACGTGACCGAGAACCTCAAATGTGGAAGAACAAACATGGTGGAGCAGATGCATGAG
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGCGTGT
ACCCGTAACTGCACCGACAAGCTGACCGCAGCACCAACAGCACCGACAGCTGGGAGAAGATG
CAACAGCACCAGCGGCCACCAACAGCACAGCACCAACAGCACCGACAGCTGGGAGAAGATG
CCCGAGGGCGAGATCAAGAACCTGCAGCTTCAAACATCACCACCGTGTGCGGACAAAGGTGCA
GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGTGCCCCTGACAAACGACAACGCCAGCT
ACCGCCTGATCAACTGCAACACCCAGCGTGTACCCAGGCCTGCCCCAAGGTGAGCTTCGAGC
CCATCCCCATCCACTACTGCGCCCCCGCCGCTTCGCCATCCTGAAGTGCAGGACAAGAAGT
TCAACGGCACCGGCCCTGCAAGAACGTGAGCACCCTGCAAGAACATCAGCACCGCATCCGCCCC
GTGGTAGCAGCCTGCTGCTGACCGCAGCCTGGCCGAGGAGGAGATCGTGTGCGCTC
CGAGAACTTACCGACAACGCCAAGACCATCATCGTCAGCTGAACGAGTCCGTGGAGATCA
ACTGCATCCGCCCCAACAAACAACACGCGTAAGAGCATCCACATCGGCCCCGGCGCCTTCT
ACGCCAACGGCGACATCATCGGCGACATCCGCCAGGGCCACTGCAACATCAGCAAGGCCAAC
TGGACCAACACCCCTCGAGCAGATCGTGGAGAACGAGCTGCGCAGGAGTTCGGCAACAACAAGAC
CATCATCTTCAACAGCAGCAGCGGGCGAACCGAGATCGTGTCCACAGCTTCAACTGCGG
CGGGAGTTCTTACTGCAACACCGAGCTGTTCAACAGCACCTGGAACATACCGAGGA
GGTAGACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGCACTCGCCAGATCATCA
ACATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCCCCCACATCCGCGGCCAGATCAAGTGC
AGCAGCAATTACCGGCCCTGCTGCTGACCCCGACGGCGGCCACCAACAACACCGCACCAA
CGACACCGAGACCTTCCGCCCGGGCGCAACATGAAGGACAACGGCAGCGAGCTGT
ACAAGTACAAGGTGGTGCATCGAGCCCCGGCGTGGCCCCACCCAGGCCAGCGCCGC
GTGGTAGCAGCGAGAACGCTAAGATATCGGATCTCTAGA

FIG. 41
(SEQ ID NO:54)

gp120.mod.US4.del128-194

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGTGG
AGCAGTCTCGTTGCCAGGCCACCACCGTGCTGTGGTGACCGTGTACTACGGCG
TGCCCGTGTGGAAGGAGGCCACCACCCACCGCTGCGCCAGCGACGCCAAGGCTTAC
AAGGCGAGGCCACAACTGTGGGCCACCCACGCCCTGCGTGCCACCGACCCCAACCC
CCAGGAGGTGAACTGACCAACGTGACCGAGAACTTCAACATGTGGAAGAACACATGG
TGGAGCAGATGCATGAGGACATCATCAGCTGTGGGACCAGAGCCTGAAGCCCTGCGTG
AAGCTGACCCCCCTGTGCGTGGGGCAGGGAACTGCGAGACCAGCGTGTACCCAGGC
CTGCCCAAGGTGAGCTCGAGCCCATCCCCATCCACTACTGCCCCCCGCCGGCTTCG
CCATCCTGAAGTGAAGGACAAGAAAGTTCAACGGCACGGCCCTGCAAGAACGTGAGC
ACCGTGCAGTGCACCCACGGCATCCGCCCGTGGTGGCAGCACCCAGCTGCTGTAACGG
CAGCCTGGCCGAGGAGGAGATCGTGTGCCCTCGAGAACTTCACCGACAACGCCAAGA
CCATCATCGTGCAGCTGAACGAGTCCGTGGAGATCAACTGCATCCGCCAACAAC
ACCGTAAAGAGCATCCACATCGGCCCCGGCGCCCTCTACGCCACCGGCACATCAT
CGGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCCACTGGACCAACACCTCG
AGCAGATCGTGGAGAAGCTGCGAGCAGTCGGCAACAAGACCATCATCTCAAC
AGCAGCAGCGGGCGGCGACCCCGAGATCGTGTCCACAGCTCAACTGCGCGGCGAGTT
CTTCTACTGCAACACCAAGCCAGCTGTTAACAGCACCTGGAACATCACCGAGGAGGTGA
ACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGCATCCGCCAGATCATCAAC
ATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCTCCATCCGCCAGATCAAGTG
CAGCAGCAATTACCGGCCCTGCTGCTGACCCGCGACGGCGCACCAACAACACCGCA
CCAACGACACCGAGACCTCCGCCCGCGCGCAACATGAAGGACAACACTGGCGCAGC
GAGCTGTACAAGTACAAGGTGGTGCATCGAGCCCTGGCGTGGCCCCACCCAGGC
CAAGCGCCGCGTGGTGCAGCGAGAAGCGCTAAGATATCGGATCCTCTAGA

FIG. 42

(SEQ ID NO:55)

55 / 131

gp140.modUS4

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGCTGTGGAGCA
GTCTTCGTTGCCGCCAGGCCACCAACCGTGCTGTGGGTGACCGTGACTACGGCGTCCCCTG
TGGAAAGGAGGCCACCAACCCCTGTTCTGCCAGCGACGCCAAGGCTTACAAGGCCAGGC
CCACAAACGTGTGGGCCACCCACGCCCTGCGTCCCCACCGACCCCAACCCCCAGGAGGTGAACC
TGACCAACGTGACCGAGAACCTCAACATGTGAAAGAACAAACATGGTGGAGCAGATGCATGAG
GACATCATCAGCCTGTGGGACCAAGCCTGAAGCCTGCGTGAAGCTGACCCCCCTGCGTG
ACCCCTGAACCTGACCGACAAGCTGACCGCAGCACCAACGGCACCAACAGCACCAAGCGGCAC
CAACAGCACCAGCGGCCACCAACAGCACCAACAGCACCGACAGCTGGGAGAAGATG
CCCGAGGGCGAGATCAAGAACCTGCAAGCTTCAACATCACCACCGCGTGCACGACAAGGTGCA
GAAGGAGTACAGCCTGTTCTACAAGCTGGACGGTGGCCATCGACAACGACAACGCCAGCT
ACCGCCTGATCAACTGCAACACCAAGCGTGAACCCAGGCGTCCCCAAGGTGAGCTCGAGC
CCATCCCCATCCAATCTCGCCTCCCGGGCTCGCCATCCTGAAGTGAAGGACAAGAAGT
TCAACGGCACCGGCCCTGCAAGAACGTGAGCACCGTGCAGTGCACCCACGGCATCCGCC
GTGGTGGACCCAGCTGCTGTAACGGCAGCCTGGCGAGGAGGAATCGTGTGCGCTC
CGAGAACATTACCGACAACGCCAACGACCATCTCGTGCAGCTGAACGAGTCCGTGGAGATCA
ACTGCATCCGCCAACACAACACCGCTGAAGAGCATCCACATCGGCCCGGCCCTCT
ACGCCACCGGCCACATCATCGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAAC
TGGACCAACACCCCTGAGCAGATCGTGGAGAACGCTGCGCGAGCAGTTCGGAACAAACAGAC
CATCATCTTCAACAGCAGCAGCGGGCGACCCGAGATCGTGTCCACAGCTTCAACTGCGG
CGCGAGTTCTTCACTGCAACACCAAGCGCAGCTGTTCAACAGCACCTGGAAACATCACCAGGA
GGTGAACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGATCCGCCAGATCATCA
ACATGTGGCAGGAGGTGGCAAGGCCATGTA CGCCCCCCCCATCCGCGGCCAGATCAAGTGC
AGCAGCAATTACCGGCCCTGCTGTCACCCCGACGGCGACCAACAAACACCGCACCAA
CGACACCGAGACCTTCCGCCCGGGCGGCAACATGAAGGACAACCTGGCGAGCGAGCTGT
ACAAGTACAAGGTGGTGCATCGAGCCCCCTGGCGTGGCCCCACCCAGGCAAGCGCCGC
GTGGTGCAGCGCGAGAACGCGCGCCGTGGCGCCCTGTTCATCGGCTCCTGGCGCC
GCCGGGAGCACCATGGCGCCGCCGTGACCTGACCGTGAGGCCAGCTGCTGAG
CGGCATCGTGCAGCAGCAGAACAAACCTGCTGCGGCCATCGAGGCCAGCAGCACCTGCTGC
AGCTGACCGTGTTGGGCATCAAGCAGCTGCAAGGCCCATCTGGCGTGGAGCGCTACCTG
AAGGACCAAGCAGCTGCTGGCATCTGGGCATCGAGCGCAAGCTGATCTGCACCAACCGT
GCCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAACATGACCTGGA
TGGAGTGGGAGCGCGAGATCGGCAACTACACCGCCCTGATCTACAACCTGATCGAGATGCC
CAGAACCAAGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCAAGTGGGCCAGCCTGT
GGAACACTGGTTCGACATACCAACTGGCTGTGGTACATCTAAGATATCGGATCCTCTAGA

FIG. 43

(SEQ ID NO:56)

gp140.mut.modUS4

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGCTGTGGAGCA
 GTCTTCGTTTGCACCAGGCCACCAACCGTGCTGTGGGTGACCGTGTACTACGGCGTCCCCGTG
 TGGAAGGAGGCCACCACCCCTGTTCTGCCAGCGACGCCAAGGCTTACAAGGCCGAGGC
 CCACAAACGTGTGGGCCACCCACGCCCTGCGTGCCCACCGACCCCAACCCCCCAGGAGGTGAACC
 TGACCAAACGTGACCGAGAACCTCAACATGTGGAAGAACAAACATGGTGGAGCAGATGCATGAG
 GACATCATCAGCCTGTGGGACCAAGGCCTGAAGGCCCTGCGTGAAAGCTGACCCCCCTGCGTG
 ACCCTGAACCTGCACCGACAAGCTGACCGGCAGCACCAACGGCACCAACAGCACCGCGCAC
 CAACAGCACCAGCGGCCACCAACAGCACCAGCACCAACAGCACCGACAGCTGGGAGAACATG
 CCCGAGGGCGAGATCAAGAACTGCAGCTTCAACATCACCACCAGCGTGCACGACAAAGGTGCA
 GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCTATCGACAACGACAACGCCAGCT
 ACCGCCTGATCAACTGCAACACCAGCGTGATCACCCAGGCCCTGCCCAAGGTGAGCTTCGAGC
 CCATCCCCATCCACTACTGCCCCCGCCGGCTCGCCATCCTGAAGTGCAAGGACAAGAAGT
 TCAACGGCACCGGCCCTGCAAGAACGTGAGCACCGTGCAGTGCACCCACGGCATCCGCCCC
 GTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCGAGGAGGAAGTCGTGCTGCCTC
 CGAGAACCTCACCGACAACGCCAACGACCATCATCGTGCAGCTGAACGAGTCCGTGGAGATCA
 ACTGCATCCGCCCCAACAAACACCGTAAGAGCATCCACATCGCCCCGGCCGCCCTCT
 ACGCCACCGCGACATCATCGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAAC
 TGGACCAACACCCCTGAGCAGATCGTGGAGAACGCTGCAGCAGTTCGGCAACAACAAGAC
 CATCATCTTCAACAGCAGCAGCGGGCGACCCCGAGATCGTGTCCACAGCTTCAACTGCAG
 CGGCGAGTTCTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGA
 GGTGAACAAGACCAAGGAGAACGACACCATCATCTGCCCTGCCATCCGCCAGATCAAGTGC
 ACATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCTCCATCCGCCAGATCAAGTGC
 AGCAGCAATTACCGGCTGCTGCTGACCCCGACGGCGGCCACAAACAAACCGCACCA
 CGACACCGAGACCTTCCGCCCGGGCGGCCACATGAAGGACAACCTGGCGCAGCAGCTGT
 ACAAGTACAAGGTGGTGCATCGAGCCCTGGCGTGGCCCCACCCAGGCCAACGCC
 GTGGTGAGCGCAGAACAGCGCCGTGGCCTGGCGCCCTGTTCATCGGCTTCTGGCGCC
 GCCGGGAGCACCATTGGCGCCGCCGTGACCCGTGACCGTGCAGGCCGCCAGCTGCTGAG
 CGGCATCGTGCAGCAGCAGAACACCTGCTGCGCCATCGAGGCCAGCAGCACCTGCTGC
 AGCTGACCGTGTGGGCATCAAGCAGCTGCAGGCCGCATCTGGCGTGGAGCGCTACCTG
 AAGGACCAAGCAGCTGGCATCTGGCGTGCAGCGCAAGCTGATCTGCACCAACCGT
 GCCCTGGAACAGCAGCTGGAGCAACAAGAGCTGACCGAGATCTGGGACAACATGACCTGGA
 TGGAGTGGGAGCGCGAGATCGGCAACTACACCGGCCATCTACAACCTGATCGAGATGCC
 CAGAACCAAGCAGGAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGT
 GGAACTGGTTGACATCACCAACTGGCTGTGGTACATCTAAGATATCGGATCCTCTAGA

FIG. 44

(SEQ ID NO:57)

gp140.TM.modUS4

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
 GTCTCGTTGCCAGGCCACCACCGTGTGGGTACCGTGACTACGGCGTGCCTG
 TGGAGGGAGGCCACCAACCTGTTCTGCCAGCGACGCCAAGGCTACAAGGCCAGGC
 CCACAAACGTGTGGCCACCAACGCCCTGCGTGCCTGCCCCACCGACCCCAACCCCAAGGAGGTGAACC
 TGACCAAACGTGACCGAGAACCTCAACATGTGGAAGAACAAACATGGTGGAGCAGATGCATGAG
 GACATCATCAGCCTGTGGGACAGAGCCTGAAGGCCCTGCGTGAAGCTGACCCCCCTGCGTG
 ACCCTGAACCTGACCGACAAGCTGACCGCAGCACCAACGGCACCAACAGCACCGACAGCTGGAGAACATG
 CAACAGCACCAGCGCACCAACAGCACCGACCAACAGCACCGACAGCTGGAGAACATG
 CCCGAGGGCGAGATCAAGAACCTGAGCTTCAACATCACCACCGCGTGCACAAAGGTGCA
 GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCTGACAAACGACAACGCCAGCT
 ACCGCTGATCAACTGCAACACCAACAGCGTGTACCCAGGCGTGCCTGACCAAGGTGAGCTTCGAGC
 CCATCCCCATCCACTACTGCCCTGCAAGAACGCTGAGCACCGTGCAGTGCACCCACGGCATCCGCC
 TCAACGGCACCGGCCCTGCAAGAACGCTGAGCACCGTGCAGTGCACCCACGGCATCCGCC
 GTGGTGAGCACCCAGCTGCTGCTGAAACGGCAGCCTGGCGAGGAGGAGATCGTGTGCGCTC
 CGAGAACCTCACCGACAACGCCAACGACCATCATCGTGCAGCTGAAACGAGTCCGTGGAGATCA
 ACTGCATCCGCCAACAACACACGCGTAAGAGCATCCACATCGGCCCGGCCGCGCTTCT
 ACGCCACCGGCCGACATCATCGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAAC
 TGGACCAACACCTCGAGCAGATCGTGGAGAACGCTGCGGAGCAGTCCGCAACAAAGAC
 CATCATCTTCAACAGCAGCAGCGGCCGACCCCGAGATCGTGTCCACAGCTTCAACTGCGG
 CGGCGAGTTCTTACTGCAACACCAACGCGTCAAGAGCATCCACATCGGCCCGGCCGCG
 GGTGAACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGATCCGCCAGATCATCA
 ACATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCGGCCATCCGCCGCGAGATCAAGTGC
 AGCAGCAATTACCGCCTGCTGCTGACCCCGCACGGCGAACAAACAAACCGCACCA
 CGACACCGAGAACCTCCGCCCGGCCGCAACATGAAGGACAACCTGGCGAGCAGCTGT
 ACAAGTACAAGGTGGTGCATCGAGGCCCTGGCGTGGCCCGACCCAGGCCAACGCC
 GTGGTGAGCGCGAGAACGCGGCCGTGGCGCCCTGTTCATCGGCTTCTGGCGCC
 GCCGGGAGCACCATGGCGCCGCCCTCGTGAACCGTGCAGGCCGAGCTGCTGAG
 CGGCATCGTGCAGCAGCACAAACCTGCTGCGGCCATCGAGGCCAGCAGCACCTGCTGC
 AGCTGACCGTGTGGGCATCAAGCAGCTGCAAGGCCATCTGGCGTGGAGCGCTACCTG
 AAGGACCAAGCAGCTGGCATCTGGGCTGCAAGCGAACAGCTGATCTGACCAACCGT
 GCCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGACAAACATGACCTGGA
 TGGAGTGGAGCGCGAGATCGGCAACTACACCGCCTGATCTACAACCTGATCGAGATCGCC
 CAGAACAGCAGGAGAACGAGCAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGT
 GGAACCTGGTTCGACATCACCAACTGGCTGGTACATCCGATCTCATGATCGTGGCG
 GCCTGATCGGCCCTGCGCATCGTGTGCGTGCAGCATCGTGAAGATATCGGATCCTCTA
 GA

FIG. 45

(SEQ ID NO:58)

Gp140modUS4.DV1V2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGC
 TGTGTGGAGCAGTCTCGTTGCCAGGCCACCACCGTCTGTGGGTGACC
 GTGTACTACGGCGTCCCCGTGAGGAGGCCACCACCCACCTCTGCG
 CCAGCGACGCCAAGGCTTACAAGGCCAGGCCACAAACGTGTGGGCCACCCA
 CGCCTGCGTGCACCAGCCCCAACCCAGGAGGTGAACCTGACCAACGTG
 ACCGAGAACCTCAACATGTGAGAACAAACATGGTGGAGCAGATGCATGAG
 GACATCATCAGCCTGTGGGACCAAGAGCCTGAAGGCCCTGCGTGGCGCCGGCC
 AGGCCTGCCCAAGGTGAGCTCGAGCCCATCCCCATCCACTACTGCGCCCC
 CGCCGGCTTCGCCATCCTGAAGTGAAGGACAAGAACAGTTCAACGGCACCGC
 CCCTGCAAGAACGTGAGCACCGTGCAGTCACCCACGGCATCCGCCCGTGG
 TGAGCACCCAGCTGCTGTAACGGCAGCCTGGCCGAGGAGGAGATCGTGCT
 GCGCTCCGAGAACCTCACCGACAACGCCAAGACCATCATCGTGCAGCTGAAC
 GAGTCCGTGGAGATCAACTGCATCCGCCAACAACAAACACCGCGTAAGAGCA
 TCCACATCGGCCCGGCCGCTTACGCCACCGGACATCATCGGCGA
 CATCCGCCAGGCCACTGCAACATCAGCAAGGCCACTGGACCAACACCTC
 GAGCAGATCGTGGAGAACAGCTGCGCGAGCAGTCCGCAACAACAAGACCATC
 ATCTTCAACAGCAGCAGCGGGCGACCCGAGATCGTGTCCACAGCTTCA
 ACTGCGGCGCGAGTTCTACTGCAACACCAGCCAGCTGTTAACAGCAC
 CTGGAACATCACCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT
 CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAGGAGGTGGCAAG
 GCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGCAGCAGCAATATTA
 CCGGCCTGCTGCTGACCCCGACGGCGCACCAACAACAACCGCACCAACGA
 CACCGAGACCTCCGCCCGCGCGCAACATGAAGGACAACGGCGCAGC
 GAGCTGTACAAGTACAAGGTGGTGCATCGAGCCCCGGCGTGGCCCCCA
 CCCAGGCCAAGCGCCGCGTGGTGCAGCGCGAGAACGGCGCCGTGGCCTGG
 GCGCCCTGTTCATCGCTCTGGCGCCGCCAGCAGCACCTGCTGCAGC
 CTCCGTACCGCTGACCGTGCAGGCCAGCTGAGGCCAGCAGCACCTGCTGCAGC
 CAGCAGAACACCTGCTGCGCGCATCGAGGCCAGCAGCACCTGCTGCAGC
 TGACCGTGTGGGCATCAAGCAGCTGCAGGCCAGCTGGCATCTGGCGTGGAGCG
 CTACCTGAAGGACCAAGCAGCTGCTGGCATCTGGGCTGCAGCGCAAGCTG
 ATCTGCACCAACCCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGA
 CCGAGATCTGGACAACATGACCTGGATGGAGTGGAGCGCAGATCGGCA
 ACTACACCGGCCTGATCTACAACCTGATCGAGATCGCCAGAACACCAGGA
 GAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGCCAGCCTGTGGAA
 CTGGTTGACATACCAACTGGCTGTGGTACATCTAAGATATCGGATCCTCTA
 GA

FIG. 46

(SEQ ID NO:59)

Gp140modUS4.DV2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGC
 TGTGTGGAGCAGTCTCGTTGCCAGGCCACCACCGTGTGGTGGGTGACC
 GTGTACTACGGCGTCCCCTGTGGAAGGAGGCCACCACCCCTGTTCTGCG
 CCAGCGACGCCAAGGCTACAAGGCCAGGCCACAAACGTGTGGGCCACCCA
 CGCCTGCGTGCCTGCCACCGACCCCAACCCCCAGGAGGTGAACCTGACCAACGTG
 ACCGAGAACCTCAACATGTGGAAGAACAAACATGGTGGAGCAGATGCATGAG
 GACATCATCAGCCTGTGGGACCAGAGCCTGAAGGCCCTGCGTGAAGCTGACCC
 CCCTGTGCGTGACCTGAACCTGCACCGACAAGCTGACCGGCAGCACCAACGG
 CACCAACAGCACCAGCGGCACCAACAGCACCAACAGCACCAACAGCACCAAG
 CACCAACAGCACCGACAGCTGGGAGAAGATGCCCGAGGGCGAGATCAAGAA
 CTGCACTTCACATCGGCCGCCGCCCTGATCAACTGCAACACACCAGCGTG
 ATCACCCAGGCCTGCCCAAGGTGAGCTCGAGGCCATCCCCATCCACTACT
 GCGCCCCCGCCGGCTCGCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGG
 CACCGGCCCTGCAAGAACGTGAGCACCCTGCACTGCAACCCACGGCATCCGC
 CCCGTGGTGAGCACCCAGCTGCTGTAACGGCAGCCTGGCGAGGAGGAGA
 TCGTGTGCGCTCCGAGAACCTCACCGACAACGCCAACGACATCATCGTGCA
 GCTGAACGAGTCCGTGGAGATCAACTGCATCCGCCAACAAACAACACCGCT
 AAGAGCATCCACATCGGCCGCCGCCCTGATCAACTGCAACACAGCGT
 TCGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAACGGACATCA
 CACCCCTGAGCAGATCGTGGAGAAGCTGCGCAGCAGTTCGGCAACAAACAA
 GACCATCATCTTCACAGCAGCAGCGGCCGACCCGAGATCGTGTCCAC
 AGCTTCACACTGCGGCCGAGTTCTTACTGCAACACACCAGCCAGCTGTTCAA
 CAGCACCTGGAACATCACCGAGGAGGTGAACAAAGACCAAGGAGAACGACAC
 CATCATCCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAGGAGGTG
 GGCAAGGCCATGTACGCCCTGCCGATCCGCCAGATCAAGTGCAGCAGCA
 ATATTACCGGCCTGCTGCTGACCCCGACGGCGGCCAACAAACAACCGCAC
 CAACGACACCGAGACCTCCGCCGCCGCCGCAACATGAAGGACAACCTG
 GCGCAGCGAGCTGTACAAGTACAAGGTGGTGCATCGAGGCCCTGGCGTG
 GCCCCCCACCCAGGCCAACGCCGCCGTGGTGCAGCGCAGAAGCGCGCCGTG
 GGCCTGGCGCCCTGTCATCGCTTCTGGCGCCGCCAGCTGCTGAGCGGCAT
 CGTGCAGCAGCAGAACACCTGCTGCGCGCCATCGAGGCCAGCAGCACCTG
 CTGCAGCTGACCGTGTGGGCATCAAGCAGCTGCAGGCCGATCCTGGCG
 TGGAGCGCTACCTGAAGGACCAGCAGCTGCTGGCATCTGGGCTGAGCGGG
 CAAGCTGATCTGCACCAACCACCGTGCCTGGAACAGCAGCTGGAGCAACAAG
 AGCCTGACCGAGATCTGGACAACATGACCTGGATGGAGTGGAGCGCGAG
 ATCGGCAACTACACCGGCCTGATCTACAACCTGATCGAGATGCCCAAGAAC
 AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCC
 TGTGGAACTGGTCGACATACCAACTGGCTGTGGTACATCTAAGATATCGG
 ATCCTCTAGA

FIG. 47
(SEQ ID NO:60)

Gp140modmutUS4.DV1V2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTTGCTGTGCTGCTGC
TGTGGAGCAGTCTCGTTGCCAGGCCACCACCGTCTGTGGGTGACC
GTGTACTACGGCGTCCCCGTGAGGAGGCCACCACCCCTGTTCTGCG
CCAGCGACGCCAAGGCTTACAAGGCCAGGCCACAAACGTGTGGGCCACCC
ACGCCTGCGTGCCCACCGACCCCAACCCCCAGGAGGTGAACCTGACCAACGT
GACCGAGAACCTAACATGTGGAAGAACAAACATGGTGGAGCAGATGCATGA
GGACATCATCAGCCTGTGGGACCAGAGCCTGAAGGCCCTGCGTGGCGCCGGC
CAGGCCTGCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCC
CCGCCGGCTCGCCATCCTGAAGTGAAGGACAAGAACGTTAACGGCACCGG
CCCTGCAAGAACGTGAGCACCGTGCAGTCACCCACGGCATCCGCCCGTG
GTGAGCACCCAGCTGCTGTAACGGCAGCCTGGCGAGGAGGAGATCGTGC
TGCCTCCGAGAACCTCACCGACAAACGCCAAGACCATCATCGTGCAGCTGAA
CGAGTCCGTGGAGATCAACTGCATCCGCCCCAACAAACACACCGTAAAGAGC
ATCCACATCGGCCCCGGCGCCTTCTACGCCACGGGACATCATCGCG
ACATCCGCCAGGCCACTGCAACATCAGCAAGGCCACTGGACCAACACCC
CGAGCAGATCGTGGAGAACGCTGCGCAGCAGTCCGGCAACAAAGACCAT
CATCTTCAACAGCAGCAGCGGGCGGCGACCCCGAGATCGTGTCCACAGCTC
AACTGCGGGCGAGTTCTTCTACTGCAACACCAGCCAGCTGTTCAACAGCA
CCTGGAACATCACCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCA
TCCCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAGGAGGTGGCAA
GGCCATGTACGCCCTGGCGAGATCAAGTGCAGCAGCAATATT
ACCGGCCTGCTGCTGACCCCGCAGCGGGCACCAACAACACCGCACCAACG
ACACCGAGACCTTCCGCCCGGGCAACATGAAGGACAACGGC
GCGAGCTGTACAAGTACAAGGTGGTGCATCGAGCCCTGGCGTGGCC
CACCCAGGCCAGCGCCGCTGGTGCAGCGCAGAACAGCGCCGTGGCCT
GGCGCCCTGTTCATCGGCTTCTGGCGCCGGAGCACCATGGCGCC
GCCTCCGTGACCGTGCAGGCCATCGAGGCCAGCAGCACCTGCTGCA
GCTGACCGTGTGGGCATCAAGCAGCTGCAGGCCAGCAGCACCTGCTGCA
CGCTACCTGAAGGACCAAGCAGCAGCTGCTGGCATCTGGGCTGCAGCG
TGATCTGCACCAACCACCGTGCCTGGAACAGCAGCTGGAGCAACAAGAGC
GACCGAGATCTGGACAACATGACCTGGATGGAGTGGAGCGCGAGATCG
CAACTACACCGGCCTGATCTACAACCTGATCGAGATCGCCAGAACAGCAG
GAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAAGTGGGCCAGCCTG
AACTGGTTGACATACCAACTGGCTGTGGTACATCTAAGATATCGGATCCTC
TAGA

FIG. 48

(SEQ ID NO:61)

gp140.mod.US4.del128-194

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTTGCTGTGCTGCTGTGTGG
AGCAGTCTCGTTGCCAGGCCACCACCGTGTGGGTGACCGTGTACTACGGCG
TGCCCGTGTGGAAGGAGGCCACCACCCACCGTGTGGGTGACGCCAGCGACGCCAAGGCTTAC
AAGGCCGAGGCCACAACGTGTGGGCCACCCACGCCACCGTGTGCCACCGACCCAAACCC
CCAGGAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGTGGAAGAACAAACATGG
TGGAGGAGATGCATGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGGCCCTGCGTG
AAGCTGACCCCCCTGTGCGTGGGGCAGGGAACTGCGAGACCAGCGTGTACCCAGGC
CTGCCCAAGGTGAGCTCGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTCG
CCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCGGCCCTGCAAGAACGTGAGC
ACCGTGCAGTGACCCACGGCATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAAACGG
CAGCCTGGCCGAGGAGGAGATCGTGTGCGCTCCGAGAACCTCACCGACAACGCCAAGA
CCATCATCGTGCAGCTGAACGAGTCCGTGGAGATCAACTGCATCCGCCCAACAACAAAC
ACCGTAAGAGCATCCACATCGGCCCCGGCGCCTTACGCCACCGGACATCAT
CGGCACATCCGCCAGGCCACTGCAACATCAGCAAGGCCACTGGACCAACACCTCG
AGCAGATCGTGGAGAACGCTGCGCAGCAGTCCGCAACAACAGACCATCATCTCAAC
AGCAGCAGCGGGCGACCCGAGATCGTGTCCACAGCTTCAACTGCGGGCGAGTT
CTTCTACTGCAACACCAGCAGCTGTTCAACAGCACCTGGAACATCACCGAGGAGGTGA
ACAAGACCAAGGAGAACGACACCACATCCTGCCCTGCCGCATCCGCCAGATCATCAAC
ATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCTGCCGCAGATCAAGTG
CAGCAGCAATATTACCGCCTGCTGCTGACCCCGCAGGGCGGCAACATGAAGGACAACGGCA
CCAACGACACCGAGACCTTCCGCCCGGGCGGCAACATGAAGGACAACGGCA
GAGCTGTACAAGTACAAGGTGGTGCAGCGCAGAGCGCAGGCCGTGGCGCTGGGCCCTGTTCATCG
CAAGCGCCGCGTGGTGCAGCGCAGAGCGCAGGCCGTGGCGCTGGCGCCCTGTTCATCG
GCTTCTGGGCCGCGCCGGAGCACCATGGCGCCCTCCGTGACCCGTGAGCAG
GCCCGCCAGCTGCTGAGCGCATCGCAGCAGCAGAACACTGCTGCGGCCATCGA
GGCCCGAGCACCTGCTGAGCTGACCGTGTGGGCATCAAGCAGCTGCAAGGCCGA
TCCTGGCCGTGGAGCGTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGCTGCAGC
GGCAAGCTGATCTGCACCAACCGTCCCTGGAACAGCAGCTGGAGCAACAAGAGCCT
GACCGAGATCTGGACAACATGACCTGGATGGAGTGGAGCGCGAGATCGGCAACTACA
CCGGCCTGATCTACAACCTGATCGAGATGCCAGAACAGCAGGAGAACGAGCAG
GAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTTCGACATACCAACTG
GCTGTGGTACATCTAAGATATCGGATCCTCTAGA

FIG. 49

(SEQ ID NO:62)

62 / 131

gp140.mut.mod.US4.del128-194

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTTGCTGTGCTGCTGTGTGG
AGCAGTCTCGTTGCCAGGCCACCACCGTGTGGGTGACCGTGTACTACGGCG
TGCCCGTGTGAAAGGAGGCCACCAACCTGTTCTGCGCCAGCGACGCCAAGGCTTAC
AAGGCCGAGGCCACACGTGTGGCCACCCACGCCCTGCGTGCACCGACCCCAACCC
CCAGGAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGTGGAAGAACAAACATGG
TGGAGCAGATGCATGAGGACATCATCAGCTGTGGGACCAGACGCTGAAGCCCTGCGTG
AAGCTGACCCCCCTGTGCGTGGGGCAGGGAACTGCGAGACCAGCGTGTACCCAGGC
CTGCCCCAAGGTGAGCTCGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTCG
CCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCGGCCCTGCAAGAACGTGAGC
ACCGTGCAGTGCACCCACGGCATCCGCCCCGTGGTGAGCACCCAGCTGCTGTGAACGG
CAGCCTGGCCGAGGAGGAGATCGTGCTGCCCTCGAGAACCTCACGACAACGCCAAGA
CCATCATCGTGCAGCTGAACGAGTCCGTGGAGATCAACTGCATCCGCCAACAACAAAC
ACCGTAAGAGCATCCACATCGGCCCCGGCGCCTCTACGCCACCGGCACATCAT
CGGCGACATCCGCCAGGGCCACTGCAACATCAGCAAGCCAACGGTCAACTGGACCAACACCTCG
AGCAGATCGTGGAGAAGCTGCGCAGCAGTTCGGCAACAACAAGACCATCATCTCAAC
AGCAGCAGCGGGCGACCCCGAGATCGTGTCCACAGCTTCAACTGCGCGGCGAGTT
CTTCTACTGCAACACCAAGCCAGCTGTTCAACAGCACCTGGAACATCACCAGGGAGGTGA
ACAAGACCAAGGAGAACGACACCATCATCTGCCCTGCCGCATCCGCCAGATCATCAAC
ATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCTCCATCCGCGGCCAGATCAAGTG
CAGCAGCAATTACCGCCTGCTGCTGACCCCGACGGCGGCCACCAACAACACCGCA
CCAACGACACCGAGACCTTCCGCCCGGGCGGCAACATGAAGGACAACACTGGCGCAGC
GAGCTGTACAAGTACAAGGTGGTGCACCGAGCCCTGGCGTGGCCCCCACCAGGC
CAAGCGCCGCGTGGTGCAGCGCAGAAGAGCGCCGTGGCGCCCTGGCGCCCTGTTCATCG
GCTTCTGGGCCCGCCGGAGCACCATGGCGCCCTCCGTGACCCGTGACCGTGCAG
GCCCGCCAGCTGCTGAGCGGACATCGTGCAGCAGCAGCAACACTGCTGCGGCCATCGA
GGCCCAAGCAGCACCTGCTGCAGCTGACCGTGTGGGCATCAAGCAGCTGCAGGCCGCA
TCCTGGCCGTGGAGCGCTACCTGAAGGACCAAGCAGCTGCTGGCATCTGGGCTGCA
GGCAAGCTGATCTGCACCAACCGTGCCTGGAACAGCAGCTGGAGCAACAAGAGCCT
GACCGAGATCTGGACAACATGACCTGGATGGAGTGGAGGCCAGATCGGCAACTACA
CCGGCCTGATCTACAACCTGATCGAGATGCCCAAGCAGCAGGAGAACGAGCAG
GAGCTGCTGGAGCTGGACAAGTGGGCCAGGCTGTGGAACCTGGTCACTCACCAACTG
GCTGTGGTACATCTAAGATATCGGATCCTCTAGA

FIG. 50

(SEQ ID NO:63)

gp160.modUS4

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
 GTCTCGTTGCCAGGCCACCACCGTGTGGGTGACCGTGTACTACGGCGTGCCGTG
 TGGAGGAGGCCACCACCCCTGTCGCCAGCGACGCCAAGGCTAACAGGCCAGGC
 CCACAAACGTGTGGCCACCCACGCCGCGTGCCACCGACCCCAACCCCCAGGAGGTGAACC
 TGACCAACGTGACCGAGAACCTCAACATGTGGAAGAACAAACATGGTGGAGCAGATGCATGAG
 GACATCATCAGCCTGTGGGACAGAGCCTGAAGGCCGTGCGTAAGCTGACCCCCCTGTGCGTG
 ACCCTGAACTGCACCGACAAGCTGACCGCAGCACCAACGGCACCAACAGCACCAGCGGCAC
 CAACAGCACCAGCGCACCAACAGCACCAACAGCACCGACAGCTGGAGAAGATG
 CCCAGGGCGAGATCAAGAACTGCAGCTAACATCACCAACCAGCGTGCACGACAAGGTGCA
 GAAGGGAGTACAGCCTGTTCAAAAGCTGGACGTGGTGCCTCGACAAACGACAACGCCAGCT
 ACCGCCTGATCAACTGCAACACCAGCGTGTACCCAGGCCCTGCCAACGGTGAAGCTTCGAGC
 CCATCCCCATCCACTACTGCGCCCCCGCCGGCTCGCCATCGTGAAGTGAAGGACAAGAAGT
 TCAACGGCACCGGCCCCCTGCAAGAACGTGAGCACCGTGCAGTGCACCCACGGCATCCGCCCC
 GTGGTGAAGCACCCAGCTGCTGAACGGCAGCCTGGCGAGGAGGAGATCGTGTGCGCTC
 CGAGAACTCACCGACAACGCCAACGACCATCATCGTGCAGCTGAACGAGTCCGTGGAGATCA
 ACTGCATCCGCCAACACAACACCGTAAGAGCATCCACATCGGCCCCGGCGCCCTTCT
 ACGCCACCGGCCGACATCATCGGCACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAAC
 TGGACCAACACCCCTCGAGCAGATCGTGGAGAACGCTGCGCGAGCAGTTCGGCAACAAACAGAC
 CATCATCTTCAACAGCAGCAGCGCCGAGACCCGAGATCGTGTCCACAGCTTCAACTGCGG
 CGCGAGTTCTTACTGCAACACCAGCCAGCTGTTAACAGCACCTGGAACATCACCGAGGA
 GGTGAACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGATCCGCCAGATCATCA
 ACATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCCCCCCTCCGCCAGATCAAGTGC
 AGCAGCAATTACCGGCTGCTGTCGACCCCGACGGCGACCAACAAACACCACCAA
 CGACACCGAGACCTTCCGCCCGGCCGGCAACATGAAGGAAACTGGCGAGCGAGCTGT
 ACAAGTACAAGGTGGTGCACATCGAGCCCCCTGGCGTGGCCCCCAGGCCAACGCCGC
 GTGGTGCAGCGCGAGAACGCGGCCGTGGCTGGCGCCCTGTTCATCGGCTTCTGGCGCC
 GCCGGAGCACCATGGCGCCGCTCGTGAACCTGACCGTGCAGGCCGCCAGCTGCTGAG
 CGGCATCGCAGCAGCAGAACACACTGCTCGCGCCATCGAGGCCAGCAGCACCTGCTGC
 AGCTGACCGTGTGGGCATCAAGCAGCTGCAGGCCATCTGGCGTGGAGCGCTACCTG
 AAGGACCAAGCAGCTGCTGGCATCTGGCGTGCAGCGCAAGCTGATCTGCACCAACCGT
 GCCCTGGAACAGCAGCTGGAGCAACAAAGACCTGACCGAGATCTGGGACAACATGACCTGGA
 TGGAGTGGAGCGCGAGATCGGCAACTACACCGGCTGATCTACAAACCTGATCGAGATCGCC
 CAGAACCGAGCAGGAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGT
 GGAACCTGGTTCGACATCACCAACTGGCTGTGGTACATCCGCATCTTCACTGATCGTGGCG
 GCCTGATCGGCTCGCAGTGTGTTGCCGTGCTGAGCATCGTAACCGCGTGCACGCCAGGGCT
 ACAGCCCCATCAGCTGCAACGCCCTGCCGCCAGCGCGGCCAGGCCAGCAACGCCCTGGTGCACGCCCTGCT
 ATCGAGGAGGAGGGCGGCCAGCGCGAGCGCGACCGCAGCAACGCCCTGGTGCACGCCCTGCT
 GGCCCTGATCGGACCTGCGCAGCCTGCGAGCCTGCGTGCAGCTGCTGTTCACTACCAACGCCCTGCGACCT
 GCTGCTGATCGGCCCCGACATCGTGGAGCTGCTGGGCCGCCGGCTGGAGGGCCCTGAAGT
 ACTGGTGGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAGCGCCGTGAGCCTGTT
 AACGCCACCGCCATCGCCGTGGCCGAGGGCACCGACCCGACATCGAGATCGCAGCGCAT
 CTTCCCGCCGTGATCCACATCCCCGCCGACATCGGCCAGGGCCTGGAGCGCGCCCTGCTGTA
 AGATATCGGATCCTCTAGA

FIG. 51

(SEQ ID NO:64)

gp160.modUS4.delV1

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
 GTCTTCGTTGCCAACGCCACCACCGTGTGGGTGACCGTGACTACGGCGTCCCCGTG
 TGGAAAGGAGGCCACCACCAACCTGTTCTGCGCCAGCGACGCCAAGGCTTACAAGGCCAGGC
 CCACAACTGTGTGGGCCACCCACGCCTGCGTGCCACCAGACCCCAACCCCCAGGAGGTGAACC
 TGACCAACGTGACCGAGAACTTCAACATGTGGAAGAACAAACATGGTGGAGCAGATGCATGAG
 GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGCGTG
 ACCCTGAACGTGACCGACAAGCTGGCGCCGGCGAGATCAAGAACACTGCAGCTTCAACAT
 CACCAACAGCGTGCAGCAAGGTGCAAGAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGG
 TGCCCATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCAACACCCAGCGTGTACACC
 AGGCCTGCCCAAGGTGAGCTTCGAGCCCATTCCATCACTACTGCGCCCCGCCGGCTTCG
 CCATCCTGAAGTGCAGGACAAGAACAGTCAACGGCACCGGCCCTGCAAGAACGTGAGCACC
 GTGCAGTGCACCCACGGCATCCGCCGTGGTGAACGACCCAGCTGCTGTAACGGCAGCCTG
 GCCGAGGAGGAGATCGTGCCTCGAGAACCTTACCGACAACGCCAACGACATCGT
 GCAGCTGAACGAGTCCGTGGAGATCAACTGCAACTCGCCACCGGCCAACACACCGCTAACAGCA
 TCCACATCGGCCCGGCCGCTTCAACGGCACCGGCCACATCGGCCACATCGGCCAGG
 CCCACTGCAACATCAGCAAGGCCAACGGACAAACACCCCTGAGCAGATCGTGGAGAACAGT
 CGCGAGCAGTTCGGCAACAACAAGACCATCATCTTCAACAGCAGCAGCGGGCGAACCGA
 GATCGTGTTCACAGCTCAACTGCGGCCGGAGTCTCTACTGCAACACCCAGCTGTT
 CAACAGCACCTGGAACATACCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCATCC
 TGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAGGAGGTGGCAAGGCCATGTACGCC
 CCCCCCATCCGCCGGCCAGATCAAGTGCAGCAGCAATATTACCGCCTGCTGACCCGCCAC
 GGCGGCCACCAACAACACCGCACCAACGACACCGAGACCTTCCGCCCGGGCGAACAT
 GAAGGACAACCTGGCGAGCAGCTGTACAAGTACAAGTGGTGCAGTACGAGCCCTGGGCG
 TGGCCCCCACCAGGCCAACGCCGGCGTGGTGCAGCGAGAACGCGCCGTGGCCTGGGCG
 GCCCTGTTCATCGGCTTCTGGCGCCGGAGCACCATGGCGCCGCTCCGTGACCTG
 ACCGTGCAGGCCGCCAGCTGCTGAGCGGCATCGTCAGCAGCAGAACAACTGCTGCGC
 CATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGGCATCAAGCAGCTGCAGGCC
 GCATCCTGCCGTGGAGCGCTACCTGAAGGACCAAGCAGCAGCTGCTGGCATTGGGCTGCAGC
 GGCAAGCTGATCTGCACCAACCCCGTGCCCTGGAACAGCAGCAGCTGGAGCAACAAGAGCCTGAC
 CGAGATCTGGGACAACATGACCTGGATGGAGTGGAGCGCGAGATCGGCAACTACACCGGC
 TGATCTACAACCTGATCGAGATCGCCAGAACCCAGCAGGAGAACGAGCAGGAGCTGCTG
 GAGCTGGACAAGTGGCCAGCCTGTGGAACTGGTTCGACATACCAACTGGCTGTGGTACATC
 CGCATCTTCACTGATCGTGGCGCCCTGATCGGCCTGCGCATCGTGTGCGTGTGAGC
 ATCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCACGCTGCAAGACCCGCTGCCGCCAG
 CGCGGCCCGACCGCCCCAGGGCATCGAGGAGGAGGGCGAGCGCAGCCTGTGCCGT
 GCAACCGCCTGGTGCACGCCGTGGCCCTGATCTGGGACGACCTGCGCAGCCTGTGCCGT
 TCAGCTACCAACCGCCTGCGCACCTGCTGCTGATCGTGGCCCGCATCGTGGAGCTGCTGGGCG
 GCCGCCGGCTGGAGGCCCTGAAGTACTGGTGGAACCTGCTGCAGTACTGGAGCCAGGAGCTG
 AAGAGCAGCGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCCAGGGCACCGACCG
 CATCATCGAGATCGTGCAGCGCATCTCCGCGCCGTGATCCACATCCCCCGCCGATCCGCA
 GGGCTGGAGCGCGCCCTGCTGTAAGATATCGGATCCTCTAGA

FIG. 52
(SEQ ID NO:65)

65 / 131

gp160.mod.US4.delV2

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTTGCTGTGCTGCTGTGTGG
 AGCAGTCTCGTTGCCAGGCCACCACCGTGTGGGTGACCGTGTACTACGGCG
 TGCCCGTGTGGAAGGAGGCCACCACCACTGTCGCCAGCGACGCCAAGGCTTAC
 AAGGCCGAGGCCACAACTGACCGAGAACATTCAACATGTGGAAGAACAAACATGG
 CCAGGAGGTGAACCTGACCAACGTGACCGAGAACATTCAACATGTGGAAGAACAAACATGG
 TGGAGCAGATGCATGAGGACATCATCAGCTGTGGGACCAGAGCCTGAAGCCCTGCGTG
 AAGCTGACCCCCCTGTGCGTACCCCTGAACCTGACCCGACAAGCTGACCGGCAGCACCAA
 CGGCACCAACAGCACCAGCGGCCAACACAGCACCAGCGGCCAACACAGCACCA
 ACAGCACCAGCTGGGAGAACATGCCGAGGGCAGATCAAGAACTGCAGCTCAAC
 ATCGGGCGCCGGCGCCTGATCAACTGCAACACCAGCGTGTACCCAGGCCCTGCCAA
 GGTGAGCTTCGAGCCCATTCCACTACTGCGCCCCCGCCGCTTCGCCATCCTGA
 AGTGCAAGGACAAGAACAGTTCAACGGCACCGGCCCCCTGCAAGAACGTGAGCACCGTGCAG
 TGCACCCACGGCATCCGGGGCTGGTGGAGCACCCAGCTGCTGTAACGGCAGCCTGGC
 CGAGGAGGAGATCGTGTGCGCTCCGAGAACATTCAACGACAAGCCAAGACCATCATCG
 TGCAGCTGAACGAGTCCGTGGAGATCAACTGCATCCGCCCCAACAAACACAGCGTAAG
 AGCATCCACATCGGCCCCGGCGCCTTACGCCACCGGCAGATCATCGGCAGCAT
 CCGCCAGGCCACTGCAACATCAGCAAGGCCACTGGACCAACACCTCGAGCAGATCG
 TGGAGAAGCTGCGCAGGAGCTCGGAACAAAGACCATCATCTTCAACAGCAGCAGC
 GGCGCGACCCCGAGATCGTGTCCACAGCTCAACTGCGCGCGAGTTCTTACTG
 CAACACCAGCCAGCTGTCACAGCACCTGGAACATCACCGAGGAGGTGAACAAAGACCA
 AGGAGAACGACACCATCATCCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG
 GAGGTGGGCAAGGCCATGTCAGCCCCCCCCTCGCGCCAGATCAAGTGCAGCAGCAA
 TATTACCGGCTGCTGTCACCCCGCACCGCACCAACAACACCGCACCAACGACA
 CCGAGACCTTCGCCCCGGCGCGCAACATGAAGGACAACCTGGCGCAGCGAGCTGTAC
 AAGTACAAGGTGGTGCATCGAGCCCCCTGGCGTGGCCCCCACCGGCCAGCG
 CGTGGTGCAGCGCAGAACGCGCCGTGGCGCCCTGTTCATCGGCTTCTGG
 GCGCCGCCGGAGCACCATTGGCGCCGCCCTGTGACCCCTGACCGTGCAGGCCAG
 CTGCTGAGCGCATCGTCAGCAGCAGAACACTGCTGCGGCCATCGAGGCCAGCA
 GCACCTGCTGCAGCTGACCGTGTGGGCATCAAGCAGCTGCAAGGCCATCTGGCG
 TGGAGCGCTACCTGAAGGACCAAGCAGCAGCTGCTGGCATCTGGGCTGCAAGCGCAAGCTG
 ATCTGCACCAACACCAGTGGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGAT
 CTGGACAAACATGACCTGGATGGAGTGGAGCGCGAGATCGCAACTACACCGCCCTGA
 TCTACAACCTGATCGAGATGCCAGAACAGCAGGAGAAGAACGAGCAGGAGCTGCTG
 GAGCTGGACAAGTGGCCAGCCTGTGGAACCTGGTTCGACATACCAACTGGCTGTG
 CATCCGCATCTTCATCATGATCGTGGCGCCTGATCGGCCCTGCGCATCGTGTGCG
 TGCTGAGCATCGTAACCGCGTGCACGGCCAGGGCTACAGCCCCATCAGCTGCAGACCG
 CTGCCGCCAGCGCGCCCCGACCGCCCCGAGGGCATCGAGGAGGAGGGCGGAGCG
 CGACCGCGACCGCAGCAACCGCCTGGTGCACGGCCCTGCTGGCCCTGATCTGGAGCG
 TGCACGCGCTGTGCCTGTTCAAGCTACCCGCTGCCGACCTGCTGCTGATCGTGG
 CGCATCGTGGAGCTGCTGGGCCGCCGCTGGAGGGCCCTGAAGTACTGGTGGAACCT
 GCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAGCGCCGTGAGCCTGTTCAACGCCACCG
 CCATGCCGTGGCCGAGGGCACCGACCGCATCGAGATCGCAGCGCATCTCCGC
 GCCGTGATCCACATCCCCGCCATCCGCCAGGGCTGGAGCGCGCCCTGCTGTAAGA
 TATCGGATCCTCTAGA

FIG. 53

(SEQ ID NO:66)

gp160.modUS4delV1/2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGCTGTGGAGCA
 GTCTTCGTTTCGCCAGGCCACCACCGTGTGGGTGACCGTGTACTACGGCGTCCCCGTG
 TGGAGGAGGCCACCACCAACCTGTTCTGCCAGCGACGCCAAGGCTTACAAGGCCAGGC
 CCACAACTGTGTGGGCCACCCACGCCGTGCTGCCACCGAACCCAAACCCCAAGGAGGTGAACC
 TGACCAACGTGACCGAGAACTCAACATGTGGAAGAACAAACATGGTGGAGCAGATGCATGAG
 GACATCATCAGCTGTGGGACCAGAGCCTGAAGCCCTGCGTGGGCCGCCAGGCCCTGCC
 CAAGGTGAGCTTCGAGCCCACCCCCATCCACTACTGCCGCCCCGCCGGCTCGCCATCCTGAA
 GTGCAAGGACAAGAACGTAACGGCACCGGCCCTGCAAGAACGTGAGCACCGTGAGTGCA
 CCCACGGCATCCGCCCCGTGGTGAACCCAGCTGCTGTAACGGAGCCTGGCGAGGAG
 GAGATCGTGTGCGCTCCGAGAACTTCACCGACAACGCCAAGACCATCATCGTCAGCTGAA
 CGAGTCCTGGAGATCAACTGCATCCGCCCCAACAACAAACACGCCGTAAGAGCATCCACATCG
 GCCCGGCCGCCCTCTACGCCACCGGCCGACATCATCGCGACATCGGCCAGGCCACTGCA
 ACATCAGCAAGGCCAACTGGACCAACACCCCTGAGCAGATCGTGGAGAACGCTGCGAGCAG
 TTCGGCAACAAGACCATCATCTTCAACAGCAGCAGCGGCCGAGCCGAGATCGTGT
 CCACAGCTCAACTCGGGCGAGTTCTTACTGCAACACCAGCCAAGCTGTTAACAGCAC
 CTGGAACATCACCGAGGGAGGTGAACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCC
 GCATCCGCCAGATCATCAACATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCTGCCATC
 CGCGGCCAGATCAAGTCAGCAGCAATTACCGGCTGCTGCTGACCCGCCAGGCCAC
 CAACAAACAACCGCACCAACGACACCGAGACCTCCGCCGCCGGCAACATGAAGGACA
 ACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCATCGAGCCCTGGCGTGGCC
 ACCCAGGCCAAGCGCCGCTGGTGCAGCGAGAACGCCGCTGGCGTGGCGCCCTGTT
 CATCGGCTTCTGGCGCCGCCGGAGCACCATGGCGCCCTCCGTGACCCCTGACCGTGCA
 GGCGGCCAGCTGCTGAGCGGCATCGTCAGCAGCACAAACCTGCTGCGGCCATCGAGG
 CCCAGCAGCACCTGCTGCACTGACCGTGTGGGCATCAAGCAGCTGCAGGCCGATCCTG
 GCCGTGGAGCGCTACCTGAAAGGACCAAGCAGCTGCTGGCATCTGGGCTGCGAGGAGATCT
 GATCTGCACCACCAACCGTGCCTGGAAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGATCT
 GGGACAAACATGACCTGGATGGAGTGGAGCGCGAGATCGGCAACTACACCCGCTGATCTAC
 AACCTGATCGAGATCGCCAGAACACAGCAGGAGAACGAGCAGGAGCTGCTGGAGCTGG
 ACAAGTGGGCCAGCCTGTGGAACTGGTGTGACATACCAACTGGCTGTGGTACATCCGATCT
 TCATCATGATCGTGGCGCCCTGATCGGCTGCGCATCGTGTGCGCTGCTGAGCATCGTGA
 ACCCGCTGCGCCAGGGCTACAGCCCCATCAGCCTGCGAGACCCGCTGCCGCCAGGCC
 CCCGACCGCCCCGAGGGCATCGAGGAGGAGGGCGAGCGCAGCCGAGCGAACCAACC
 GCCTGGTGCACGGCTGCTGGCCCTGATCTGGAGCACCTGCGCAGCCTGCGCTGCTGAGCT
 ACCACCGCTGCGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAGCTGCTGGGCCGCC
 GCTGGAGGCCCTGAAGTACTGGTGGAACTGCTGCACTGAGACTGGAGGCCAGGAGCTGAAGAGC
 AGCGCCGTGAGCCTGTTCAACGCCACCGCATCGCCGTGGCCAGGGCACCGACCGCATCATC
 GAGATCGTGCAGCGCATCTCCGCGCCGTGATCCACATCCCCGCCGATCCGCCAGGGCTG
 GAGCGCGCCCTGCTGTAAGATATCGGATCCTCTAGA

FIG. 54

(SEQ ID NO:67)

gp160.modUS4 del 128-194

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGTGGAGCA
 GTCTTCGTTTGCAGGCCACCACCGTCTGCTGGGTGACCGTGTACTACGGCGTCCCCGTG
 TGGAGGGAGGCCACCACCCACCTGTTCTGCGCCAGCGACGCCAAGGCTACAAGGCCAGGC
 CCACAACTGTGGGCCACCCACGCCCTGCGTCCCCACCGACCCCAACCCCCAGGAGGTGAACC
 TGACCAACGTGACCGAGAACTTCAACATGTGGAAGAACAAACATGGTGGAGCAGATGCATGAG
 GACATCATCAGCCTGTGGGACCAGAGCCTGAAGGCCCTGCGTGAAGCTGACCCCCCTGTGCGTG
 GGGGCAGGGAACTGCGAGACCAGCGTGTACCCAGGCCCTGCCCAAGGTGAGCTTCGAGCC
 CATCCCCATCCACTACTGCGCCCCCGCCGGCTCGCCATCTGAAGTGAAGGACAAGAACGTT
 CAACGGCACCGGCCCTGCAAGAACGTGAGCACCGTGCAGTGCACCCACGGCATCCGCCCG
 TGGTGAGCACCCAGCTGCTGTAACGGCAGGCCCTGGCCAGGGAGATCGTGCTGCGCTCC
 GAGAACCTTACCGACAACGCCAACGACCATCATCGTGCAGCTGAACCGAGTCCGTGGAGATCAA
 CTGCATCCGCCCAACAAACAAACCGCGTAAGAGCATCCACATCGGCCCGGCCGCGCCCTCTA
 CGCCACCGGCACATCATCGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAACT
 GGACCAACACCCCTGAGCAGATCGTGGAGAACGACTGCGCAGCAGTCCGCAACAAACAGACC
 ATCATCTTCAACAGCAGCAGCGCCGAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGAG
 GGCAGGTTCTTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGAG
 GTGAACAAGACCAAGGAGAACGACACCATCATCTGCCCTGCCGATCCGCCAGATCATCAA
 CATGTGGCAGGAGGTGGCAAGGCCATGTACGCCCTGCCGATCCGCCAGATCAAGTGCA
 GCAGCAATTACCGGCCTGCTGACCCCGCAGGCCAGCAACAAACACCGCACCAAC
 GACACCGAGACCTTCCGCCCGGCCGGCAACATGAAGGACAACCTGGCGCAGCGAGCTGTA
 CAAGTACAAGGTGGTGCACATGAGGCCCTGGCGTGGCCCCCACCCAGGCCAAGGCCG
 TGGTGAGCGCAGAGCGCGCCGTGGCCTGGCGCCCTGTTCATCGGCTTCTGGCGCCG
 CCGGGAGCACCATGGCGCCCTCCGTGACCCCTGACCGTGCAGGCCCGCAGCTGAGC
 GGCATCGTCAGCAGCAACACCTGCTGCCGCCATCGAGGCCAGCAGCACCTGCTGCA
 GCTGACCGTGTGGGCATCAAGCAGCTGCAGGCCCGCATCTGGCGTGGAGCGCTACCTGA
 AGGACCAAGCAGCTGCTGGCATCTGGGCTGACCGGCAAGCTGATCTGCACCAACCGCTG
 CCCTGGAAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGACAACATGACCTGGAT
 GGAGTGGAGCGCAGATCGCAACTACACCGGCCTGATCTACAACCTGATCGAGATCGCCC
 AGAACCAAGCAGGAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGCCAGCTGTG
 GAACTGGTTCGACATACCAACTGGCTGTGGTACATCCGATCTTATCATGATCGTGGCGG
 CCTGATCGGCCTGCGCATCGTGTGTCGCCGTGAGCATCGTGAACCGCGTGCAGGCCAGGGCTA
 CAGCCCCATCAGCCTGCAAGACCCGCTGCCGCCAGCGCGCCCCGACCGCCCCGAGGGCA
 TCGAGGAGGGCGCGAGCGCAGCCGACCGCAGCAACGCCCTGGTGACGGCCTGCTG
 GCCCTGATCTGGGACGACTGCGCAGCCTGCGCTGTTCACTGCTACCCGCCGCGACCTG
 CTGCTGATCGTGGCCCGCATCGTGGAGCTGCTGGGCCGCCGGCTGGAGGCCCTGAAAGTAC
 TGGTGGAACCTGCTGCAGTACTGGAGGCCAGGAGCTGAAAGAGCAGCGCCGTGAGCCTGTTCAA
 CGCCACCGCCATCGCCGTGGCCGAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCTT
 CGCGCCGTGATCCACATCCCCGCCGATCCGCCAGGCCCTGGAGCGCGCCCTGCTGTAAGA
 TATCGGATCCTCTAGA

FIG. 55

(SEQ ID NO:68)

68 / 131

Env_US4_C4wt

GACACTATCATACTCCCAGCAGAATAAGACAAATTATAAACATGTGGCAAGAAGTAGG
AAAAGCAATGTATGCCCTCCCATCAGAGGACAAATTAAATGTCATCAAATATTACAG
GGCTGCTATTAACAGAGATGGTGGT

FIG. 56
(SEQ ID NO:69)

Env_SF162_C4wt

GGAACTATCACACTCCCAGCAGAATAAAACAAATTATAAACAGGTGGCAGGAAGTAGG
AAAAGCAATGTATGCCCTCCCATCAGAGGACAAATTAGATGTCATCAAATATTACAG
GACTGCTATTAACAAGAGATGGTGGT

FIG. 57
(SEQ ID NO:70)

Env_US4_C4mod

GACACCATCATCCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAGGAGGTGGG
CAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGCAGCAGCAACATCACCG
GCCTGCTGCTGACCCGCGACGGCGGC

FIG. 58
(SEQ ID NO:71)

Env_SF162_C4mod

GGCACCATCACCTGCCCTGCCGCATCAAGCAGATCATCAACCGCTGGCAGGAGGTGGG
CAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCCGCTGCAGCAGCAACATCACCG
GCCTGCTGCTGACCCGCGACGGCGGC

FIG. 59
(SEQ ID NO:72)

69 / 131

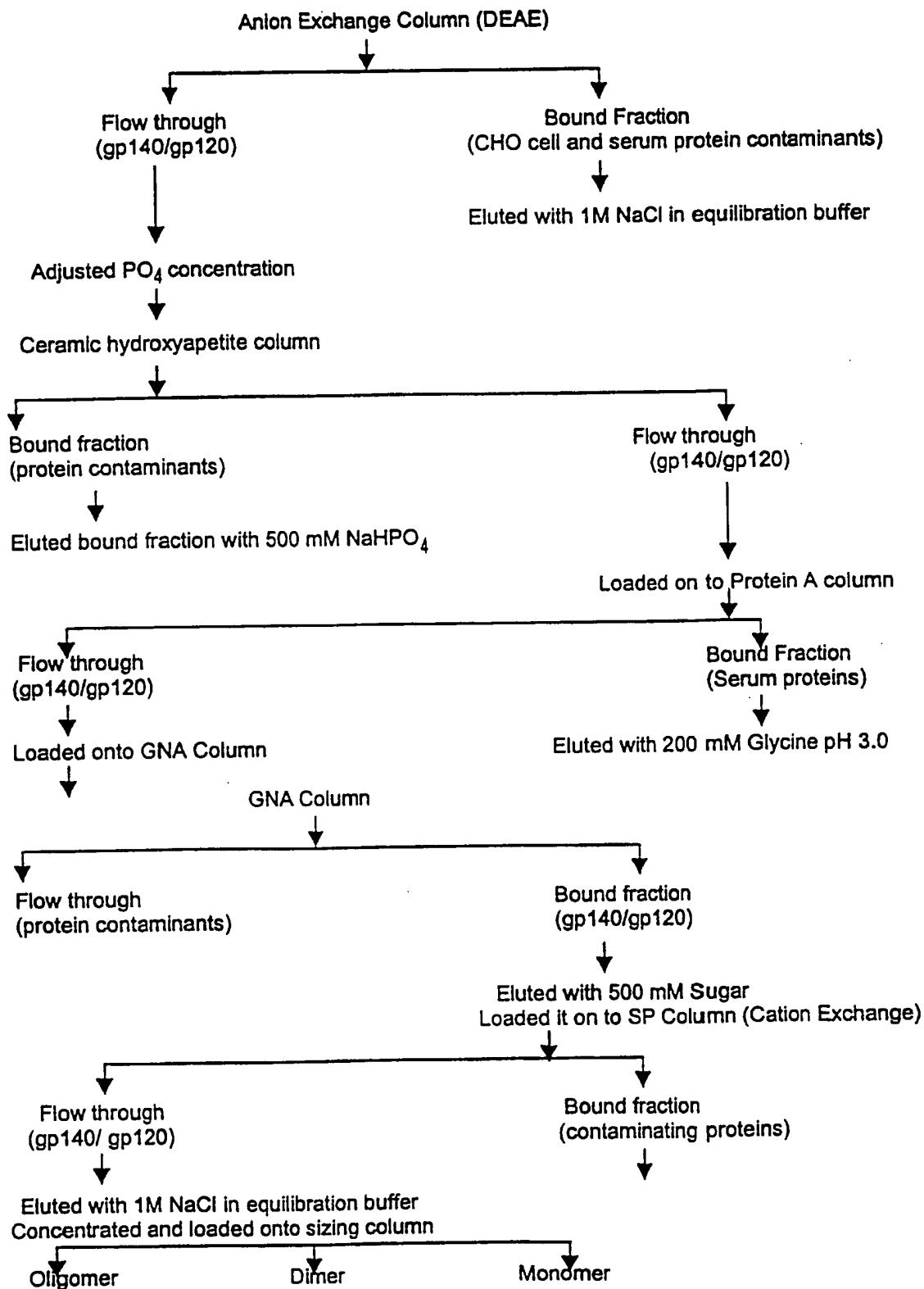


FIG. 60

70 / 131

gp160mod.us4.gag.modSF2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGCTGTGGA
 GCAGTCTTCGTTGCCAGGCCACCACCGTGCTGGGTGACCGTGTACTACGGCGTG
 CCCGTGTGGAAGGAGGCCACCACCCCTGTTCTGCCAGGCCAAGGCTTACAAG
 GCCGAGGCCACAACGTGTGGGCCACCCACGCCGTGCCCACCGACCCAAACCCCCAG
 GAGGTGAACCTGACCAACGTGACCGAGAACCTCAACATGTGGAAGAACAAACATGGTGGAG
 CAGATGCATGAGGACATCATCAGCCTGTGGGACCAAGGCCTGAAGGCCCTGCGTGAAGCTG
 ACCCCCCCTGTGCGTGACCCCTGAACTGCACCGACAAGCTGACCGGCAGCACCAACGGCACC
 AACAGCACCAGCGGCCACCAACAGCACCAGCGGCCACCAACAGCACCAACAGCAC
 GACAGCTGGGAGAAGATGCCGAGGGCGAGATCAAGAACCTGCAGCTTCAACATCACCACC
 AGCGTGCAGACAAGGTGAGAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGC
 ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCAACACACCAGCGTGTACCCAG
 GCCTGCCCAAGGTGAGCTCGAGCCCATTCCATCCACTACTGCGCCCCCGCCGGCTTC
 GCCATCCTGAAAGTGCAGGACAAGAACGTTCAACGGCACCGGCCCTGCAAGAACGTGAGC
 ACCGTGCAGTGCACCCACGGCATCCGCCCGTGGTGAGCACCCAGCTGCTGTAACGGC
 AGCCTGGCCGAGGAGGAGATGTGCTGCGCTCCGAGAACCTCACCGACAACGCCAAGACC
 ATCATCGTGCAGCTGAAAGTAGCTCGTGGAGATCAACTGCATCCGCCAACAAACAAACAG
 CGTAAGAGCATCCACATCGGCCCGGCCCTACGCCACCGGCAGCATCATCGGC
 GACATCCGCCAGGCCACTGCAACATCAGCAAGGCCACTGGACCAACACCCCTGAGCAG
 ATCGTGGAGAAGCTGCGCAGCAGTTCGGCAACAACAAGACCATCATCTTCAACAGCAGC
 AGCGCGCGGACCCCGAGATCGTGTCCACAGCTCAACTGCGCGGGCAGTTCTTCTAC
 TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGAGGTGAACAAGACC
 AAGGAGAACGACACCATCATCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG
 GAGGTGGCAAGGCCATGTACGCCCTGCCGATCCGCCAGATCAAGTGCAGCAGCAAT
 ATTACCGCCCTGCTGACCCCGACGGCGCACCAACAACACCGCACCAACCGACACC
 GAGACCTTCCGCCCGGGCGGCAACATGAAGGACAACGGCGCAGCGAGCTGTACAAG
 TACAAGGTGGTGCATCGAGCCCTGGCGTGGCCCCCACCAGGCCAGCGCCCGCTG
 GTGCAGCGCAGAACGCGCCGTGGCGCCATCCGCCAGATCATCGGCTTCTGGCGCC
 GCCGGAGCACCATGGCGCCCTCCGTGACCCCTGACCGTGAGGCCAGCAGCACCTG
 AGCGGCATCGCAGCAGCACAAACCTGCTGCGCCATCGAGGCCAGCAGCACCTG
 CTGCAGCTGACCGTGTGGGCATCAAGCAGCTGCGAGGCCGATCTGCCGTGGAGCGC
 TACCTGAAGGACCAAGCAGCTGCTGGCATCTGGGCTGCAAGGGCAAGCTGACCGAGATCTGGACAAC
 ACCACCGTGCCCTGGAACAGCAGCTGGAGAACAGCAGCTGACCGAGATCTGGACAAC
 ATGACCTGGATGGAGTGGAGCGCGAGATGGCAACTACACCGCCTGATCTACAACCTG
 ATCGAGATGCCCAAGAACAGCAGGAGAACAGCAGGAGCTGCTGGAGCTGGACAAG
 TGGGCCAGCCTGGAACCTGGTCGACATCCAACCTGGCTGTTGACATCCGCACTTC
 ATCATGATCGTGGCGGCCTGATCGGCCCTGCGCATCGTGTGCGCTGAGCATCGTG
 AACCGCGTGCCTGCCAGGGCTACAGCCCCATCAGCCTGCAAGACCCGCTGCCCGCAGCGC
 GGCCCCGACGCCCGAGGGCATCGAGGAGGGAGGGCGAGCGCGACCGCACCGCAGC
 AACCGCCTGGTCAGGCCCTGCGGACCTGCTGCTGATCGTGGCCCGATCGTGGAGCTGCTG
 TTCAGCTACCACCGCCTGCGGACCTGCTGCTGATCGTGGCCCGATCGTGGAGCTGCTG
 GGCGCCGCGGCTGGAGGGCCTGAAAGTACTGGTGGAAACCTGCTGCAAGTACTGGAGGCCAG
 GAGCTGAAGAGCAGGCCGTGAGCCTGTTCAACGCCACCGCCATGCCGTGGCCAGGGC
 ACCGACCGCATTCGAGATCGCAGCGCATCTCCGCGCCGTGATCCACATCCCCCGC
 CGCATCCGCCAGGGCTGGAGCGCCCTGCTGTAAGATATCGGATCCTTAGAGAATTG

FIG. 61A

(SEQ ID NO:73)

CGCCCCCCCCCCCCCCCCCTCCCTCCCCCCCCCTAACGTTACTGCCGAAGCCGC
TTGGAATAAGGCCGGTGTGCGTTGTCTATGTTATTTCCACCATATTGCCGTCTTT
GGCAATGTGAGGGCCGGAAACCTGGCCCTGCTTCTGACGAGCATTCTAGGGTCTT
TCCCCCTCGCCAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTCCTCTG
GAAGCTTCTTGAAGACAAACAACGCTGTAGCGACCCCTTGAGGCAGCGGAACCCCCA
CCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATAACACTGCAAAGGCG
GCACAAACCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCTCC
TCAAGCGTATTCAACAAGGGCTGAAGGATGCCAGAAGGTACCCATTGTATGGATCT
GATCTGGGGCCTCGGTGCACATGCTTACATGTGTTAGTCGAGGTTAAAAAAACGTCTA
GGCCCCCGAACCACGGGGACGTGGTTTCTTGAACACGATAATACCATGGCGC
CCGCGCCAGCGTGTGAGCGGCGGGAGCTGGACAAGTGGGAGAAGATCCGCTGCGCCC
CGGCAGGAAGAAGAAGTACAAGCTGAAGCACATCGTGTGGCCAGCGAGCTGGAGCG
CTTCGCCGTGAACCCGGCTGCTGGAGACCAGCGAGGGCTGCCAGATCCTGGCCA
GCTGCAGCCCAGCCTGCAAGACGGCAGCGAGGAGCTCGCAGCCTGTACAACACCGTGGC
CACCTGTACTGCGTGCACCGCGCATCGACGTCAAGGACACCAAGGAGGCCCTGGAGAA
GATCGAGGAGGAGCAGAACAGTCCAAGAAGAAGGCCAGCAGGCCGCCGCCGG
CACCGGCAACAGCAGCCAGGTGAGCCAGAACTACCCATCGTCAGAACCTGCAGGGCCA
GATGGTGCACCAAGGCCATCAGCCCCCGCACCTGAACGCCCTGGTGAAGGTGGTGGAGGA
GAAGGCCTCAGCCCCGAGGTGATCCCCATGTTAGCGCCCTGAGCGAGGGGCCACCC
CCAGGACCTGAACACGATGTTGAACACCGTGGCGGCCACCAGGCCATGCAGATGCT
GAAGGAGACCATCAACGAGGAGGCCAGTGGGACCGCGTGCACCCCGTGCACGCCGG
CCCCATCGCCCCGGCCAGATGCGCGAGCCCGCGCAGCGACATGCCGGCACCAACAG
CACCTGCAGGAGCAGATCGGCTGGATGACCAACAACCCCCCATCCCCGTGGCGAGAT
CTACAAGCGTGGATCATCCTGGCCTGAACAAGATCGTGCAGGATGTACAGCCCCACCAG
CATCCTGGACATCCGCCAGGGCCCCAAGGAGCCCTTCCCGGACTACGTGGACCGCTTCTA
CAAGACCCCTGCGCGTGCAGGCCAGGACGTGAAGAACTGGATGACCGAGACCC
GCTGGTGCAGAACGCCAACCCGACTGCAAGACCATCCTGAAGGCTCTGGCCCCCGGC
CACCTGGAGGAGATGATGACCGCCTGCCAGGGCGTGGCGGCCACAGGCC
CGTGCCTGGCCAGGCGATGAGCCAGGTGACGAACCCGGCAGCATCATGATGCAGCGCG
CAACTCCGCAACCAGCGGAAGACCGTCAAGTGCCTCAACTGGCAAGGAGGGCACAC
CGCCAGGAACTGCCGCCGGGGGGCAAGAAGGGCTGCTGGCGTGCAGGCCAGGCC
CCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCCCTGGCAAGATCTGGCCAGCTA
CAAGGGCCGCCGGCAACTTCCCTGCAGAGCCGCCAGGCCACCGCCCCCGAGGA
GAGCTTCCGCTTGGCGAGGAGAAGACCAACCCAGCCAGAAGCAGGAGCCATCGACAA
GGAGCTGTACCCCTGACCGCCTGCGAGCCTGGCAACGACCCAGCAGCCAGTA
AGAATTCAAGACTCGAGCAAGTCTAGA

FIG. 61B
(SEQ ID NO:73)

gp160mod.SF162.gag.modSF2

FIG. 62A

(SEO ID NO:74)

73 / 131

GTAACTCGAGCAAGTCTAGAGAATTCCGCCCCCCCCCCCCCCCCCTCTCCCTCCCC
CCCCCTAACGTTACTGGCGAAGCCGCTTGAATAAGGCCGTGTGCCTTGCTATAT
GTTATTTCCACCATAATGCCGTCTTGGCAATGTGAGGGCCCGAAACCTGGCCCTG
TCTTCTTGACGAGCATCCTAGGGTCTTCCCCTCGCCAAGGAATGCAAGGTCTG
TTGAATGTCGTGAAGGAAGCAGTTCTCTGGAAGCTTCTTGAAGACAAACAACGTCTG
AGCACCCTTGCAAGGCAGCGAACCCCCCACCTGGCAGCGTGCCTTGCGGCCAAA
AGCCACGTGTATAAGATACACCTGCAAAGGCGCACAAACCCAGTGCCACGTTGTGAGT
TGGATAGTTGTGGAAAGAGTCAAATGGCTCTCTCAAGCGTATTCAACAAGGGCTGAA
GGATGCCAGAAGGTACCCCATTGTATGGATCTGATCTGGGCCCTCGGTGCACATGCT
TTACATGTGTTAGTCAGGTTAAAAAAACGTCTAGGCCCGAACACGGGACGTG
GTTTCTTGAACACGATAATACCATGGCGCCCGCCAGCGTGTGAGCGCG
GCGAGCTGGACAAGTGGAGAAGATCCGCCTGCGCCCCGGCGCAAGAAGAAGTACAAG
CTGAAGCACATCGTGTGGCCAGCCGAGCTGGAGCGCTTCGCCGTGAACCCCGCCT
GCTGGAGACCAGCGAGGGCTGCCAGATCCTGGCCAGCTGCAGCCAGCCTGCAGA
CCGGCAGCGAGGAGCTGCGCAGCCTGTACAACACCGTGGCCACCCCTGTACTGCGTCAC
CAGCGCATCGACGTCAAGGACACCAAGGAGGCCCTGGAGAAGATCGAGGAGGAGCAGAA
CAAGTCCAAGAAGAAGGCCAGCAGGCCCGCCGCCGGCACCGCAACAGCAGCC
AGGTGAGCCAGAACTACCCATCGTGCAGAACCTGCAGGGCAGATGGTGCACCAAGGCC
ATCAGCCCCCGCACCTGAACGCCTGGGTGAAGGTGGAGGAGAAGGCCTTCAGCCC
CGAGGTGATCCCCATGTTAGCGCCCTGAGCGAGGGGCCACCCCGCAGGACCTGAACA
CGATGTTGAACACCGTGGCGGGCACCAGGCCATGCAGATGCTGAAGGAGACCATC
AACGAGGAGGCCAGTGGGACCGCGTGCACCCCGTGCACGCCGGCCCATCGCCCC
CGGCCAGATGCGCAGGCCCGCCAGCGACATGCCGGCACCAAGCAGCAGCAGGCC
AGCAGATCGCTGGATGACCAACAACCCCCCATCCCCGTGGCGAGATCTACAAGCGG
TGGATCATCTGGGCCCTGAACAAGATCGTGCAGGATGTACAGCCCCACAGCATCCTGGA
CATCCGCCAGGGCCCCAAGGAGCCCTTCCCGACTACGTGGACCGCTTCTACAAGACCC
TGCGCGCTGAGCAGGCCAGCCAGGACGTGAAGAACCTGGATGACCGAGACCTGCTGGTG
CAGAACGCCAACCCGACTGCAAGACCATCCTGAAGGCTCTGGCCCCCGCGGCCACCC
GGAGGAGATGATGACCGCCTGCGAGGGCGTGGCGGCCACAGGCCGGCGTGC
TGGCCGAGGCAGTGGCAGGTGACGAACCCGGCAGCATCATGATGCAAGCGCGGCAAC
TTCCGCAACCAGCGGAAGACCGTCAAGTGTCTCAACTGCGGAAGGAGGGCCACACCGC
CAGGAACCGCCGGCCCCCGCAAGAACGGCTGCTGGCGTGCAGGCCAGGCCAC
AGATGAAAGGACTGCACCGAGCGCCAGGCCAACTTCTGGCAAGATCTGGCCCAGCTAC
AAGGGCCGCCCGCAACTTCTGACAGGCCGGCCCCAGGCCACCGCCCCCGAGGA
GAGCTTCCGCTTGGCGAGGAGAAGACCAACCCAGGCCAGAACAGCAGGCCAC
AGGAGCTGTACCCCCCTGACCAAGCCTGCGCAGCCTGTTGGCAACGACCCAGCAGGCCAG
TAAGAATTCAACTCGAGCAAGTCTAGA

FIG. 62B
(SEQ ID NO:74)

gp160modUS4.delV1/V2.gag.modSF2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGCTGCTGCTGTGG
 GCAGTCTCGTTGCCAGGCCACCACCGTCTGGGTGACCGTGACTACGGCGTG
 CCCGTGTGGAAGGAGGCCACCACCCCTGTCGCCCCAGCAGCCAAGGCTTACAAG
 GCCGAGGCCAACACGTGTGGCCACCCACGCCCTGCGTCCCACCGACCCAAACCC
 GAGGTGAACCTGACCAACGTGACCGAGAACATCAACATGTGGAAGAACAAACATGGTGGAG
 CAGATGCATGAGGACATCATCAGCCTGTGGGACCAAGGCTGAAGCCCTGCGTGGCGCC
 GGCCAGGCCCTGCCCAAGGTGAGCTCGAGCCCATCCCCATCCACTACTGCGCCCCGCC
 GGCTTCGCCATCCTGAAGTGAAGGACAAGAACAGTTAACGGCACCGGCCCTGCAAGAAC
 GTGAGCACCGTGCAGTGCACCCACGGCATCCGCCCGTGGTGAGCACCCAGCTGCTG
 AACGGCAGCCTGGCCGAGGAGGAGATCGTGTGCGCTCCGAGAACATCACCGACAA
 AAGACCATCATCGTCAGCTGAACGAGTCGGAGATCAACTGCATCCGCCAACAA
 AACACCGCTAACAGCATCCACATCGGCCCGGCCCTACGCCACCGCGACATC
 ATCGGCACATCCGCCAGGCCACTGCAACATCAGCAAGGCCACTGGACCAACAC
 GAGCAGATCGTGGAGAACGCTGCGCAGCAGTTGGCAACAACAAGACCATCATCTCAAC
 AGCAGCAGCGGGGAGCCAGATCGTGTCCACAGCTCAACTGCGCGGAGTT
 TTCTACTGCAACACCAGCCAGCTGTTAACAGCACCTGGAACATCACCGAGGAGGTGAAC
 AAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGATCCGCCAGATCATCAACATG
 TGGCAGGAGGTGGCAAGGCCATGTACGCCCGGCGACATCCGCGCCAGATCAAGTGCAGC
 AGCAATATTACCGGCCTGCTGCTGACCCGCGACGGCGGCCAACAAACAACCGCACCAAC
 GACACCGAGACCTCCGCCCGGCCGGCAACATGAAGGACAACCTGGCGCAGCGAGCTG
 TACAAGTACAAGGTGGTGCATCGAGCCCTGGGCGTGGGCCACCCAGGCCAAC
 CGCGTGGTGCAGCGCGAGAAGCGCGCGTGGCGCCCTGTTCATCGGCTTCCTG
 GGCGCCGCCGGAGCACCATGGGCCGCGCTCCGTGACCCGTGACCGTGCAGGCCAG
 CTGCTGAGCGGCATCGTCAGCAGCAGAACACCTGCTGCGGCCATCGAGGCCAGCAG
 CACCTGCTGAGCTGACCGTGTGGGCATCAAGCAGCTGCGAGCCGATCCGGCGTG
 GAGCGCTACCTGAAGGACCAGCAGCTGCTGGCATCTGGGCTGAGCGCAAGCTGATC
 TGCACCAACCACCGTCCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGG
 GACAACATGACCTGGATGGAGTGGAGCGCGAGATCGGCAACTACACCGCCCTGATCTAC
 AACCTGATCGAGATGCCAGAACCAAGCAGCAGGAGAACGAGCAGGAGCTGCTGGAGCTG
 GACAAGTGGGCCAGCCTGGAACCTGGTCACTGACATCACCACCTGGCTGGTACATCCGC
 ATCTTCATCATGATCGTGGCGGCCCTGATCGGCCCTGCGCATCGTGTGCGCTGAGC
 ATCGTGAACCGCGTGCAGGCCAGGGCTACAGCCCCATCAGCCTGCGAGACCCGCCCTGCC
 CAGCGCGGCCCGACCGCCCGAGGGCATCGAGGAGGGCGAGCGCAGCCGAC
 CGCAGCAACCGCCTGGTGCACGGCCTGCTGGCCCTGATCTGGACGACCTGCGCAGCCTG
 TGCCTGTTCAGCTACCAACCGCCTGCGCGACCTGCTGCTGATCGTGGCCGATCGTGGAG
 CTGCTGGGCCGCCGGCTGGAGGCCCTGAAAGTACTGGTGGAACCTGCTGAGTACTGG
 AGCCAGGAGCTGAAGAGCAGCGCCGTGAGCTGTTAACGCCACCGCCATGCCGTGGCC
 GAGGGCACCGACCGCATCGAGATCGTGCAGCGCATCTTCCGCGCCGTGATCCACATC
 CCCCGCCGATCCGCCAGGGCTGGAGCGCGCCCTGCTGTAAGATATCGGATCCTCTAGA
 GAATTCCGCCCTGGGATGG
 AGCCGCTTGGATAAGGCCGGTGTGCGTTGTCTATATGTTATTTCCACCATATTGCC
 TCTTTGGCAATGTGAGGGCCGGAAACCTGGCCCTGCTTCTGACGAGCATTCTTAGG
 GGTCTTCCCTCTGCCAAAGGAATGCAAGGTCTGTTGAATGCGTGAAGGAAGCAGTT

FIG. 63A

(SEQ ID NO:75)

CCTCTGGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCTTGCAAGGCAGCGAAC
CCCCCACCTGGCGACAGGTGCCTCTGGCCAAAAGCCACGTATAAGATAACACCTGCA
AAGGCGGCACAACCCAGTGCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCATGG
CTCTCCTCAAGCGTATTCAACAAGGGCTGAAGGATGCCAGAAGGTACCCATTGTATG
GGATCTGATCTGGGCCTCGGTGCACATGCTTACATGTGTTAGTCGAGGTTAAAAAAA
CGTCTAGGCCCCCGAACACGGGACGTGGTTTCCCTTGAAAAACACGATAATACCAT
GGCGCCCGGCCAGCGTGCAGCGCCAGCTGGACAAGTGGAGAAGATCCGCCT
GCGCCCGGGCGCAAGAAGAAGTACAAGCTGAAGCACATCGTGTGGCCAGCCGAGCT
GGAGCGCTTCGGCGTGAACCCGGCTGCTGGAGACCAGCGAGGGCTGCCAGATCCT
GGGCCAGCTGCAGCCCAGCCTGCAGACCGGAGCGAGCTGCGCAGCCTGTACAACAC
CGTGGCCACCCCTGTACTGCGTGCACCAGCGCATCGACGTCAAGGACACCAAGGAGGCCCT
GGAGAAGATCGAGGAGGAGCAGAACAGTCCAAGAAGAAGGCCAGCAGGCCGCCGC
CGCCGGCACCGCAACAGCAGCCAGGTGAGCCAGAACTACCCATCGTCAGAACCTGCA
GGGCCAGATGGTGACCAGGCATCAGCCCCCGCACCCCTGAACGCCCTGGGTGAAGGTGGT
GGAGGAGAAGGCCCTCAGCCCCGAGGTGATCCCCATGTTAGCGCCCTGAGCGAGGGCGC
CACCCCCCAGGACCTGAACACGATGTTGAACACCGTGGCGGCCACCGCCCATGCA
GATGCTGAAGGAGACCATCAACGAGGAGGCCAGATGCGCAGCCCCGCCAGCAGACATCGCCGCAC
CGCCGGCCCCATCGCCCCCGGCCAGATGCGCAGCCCCGCCAGCAGACATCGCCGCAC
CACCAAGCACCTGCAGGAGCAGATCGGCTGGATGACCAACAACCCCCCATCCCCGTGG
CGAGATCTACAAGCGGTGGATCATCCTGGCTGAACAAGATCGCGGATGTACAGCCC
CACCAAGCACCTGGACATCCGCCAGGGCCCAAGGAGCCCTCGCGACTACGTGGACCG
CTTCTACAAGACCCCTGCGCGCTGAGCAGGCCAGGACGTGAAGAACTGGATGACCGA
GACCCCTGCTGGTGAGAACGCCAACCCGACTGCAAGACCATCCTGAAGGCTCTGGCCC
CGCGGCCACCCCTGGAGGAGATGATGACCGCCTGCCAGGGCGTGGCGGCCAGCAA
GGCCCGCGTGCCTGGCGAGGCAGATGAGCCAGGTGACGAACCCGGCAGCATCATGTC
GCGCGCAACTCCGCAACCAGCGGAAGACCGTCAAGTGTCAACTGCCAGGGAGGG
CCACACCGCCAGGAACCGCCGCCCCCGCAAGAACGGCTGCTGGCGCTGCCAGCG
GGGCCACCCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTGGCAAGATCTGGCC
CAGCTACAAGGGCCGCCGGCAACTTCTGCAAGGCCAGGGAGGCCACCGCCCCCCC
CGAGGAGAGCTCCGCTTGGCGAGGAGAAGACCAACCCAGGCCAGAAGCAGGAGCCAT
CGACAAGGAGCTGTACCCCTGACCGAGCCTGCGCAGCCTGTTGGCAACGACCCAGCAG
CCAGTAAGAATTCAAGACTCGAGCAAGTCTAGA

FIG. 63B

(SEQ ID NO:75)

gp160.modSF162.delV2.gag.modSF2

GAATTGCCACCATGGATGCAATGAAGAGAGGGCTTGCTGTGCTGCTGTGTGGA
 GCAGTCTCGTTGCCAGGCCGTGGAGAAGCTGTGGGTGACCGTGTACTACGGCGTG
 CCCGTGTGGAGGAGGCCACCACCCCTGTTCTGCCAGCAGCCAAGGCCTACGAC
 ACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCCAACCCCCAG
 GAGATCGTGTGGAGAACGTGACCGAGAACCTCAACATGTGGAAGAACAAACATGGTGGAG
 CAGATGCACGAGGACATCATCAGCCTGTGGGACAGAGCCTGAAGGCCCTGCGTGAAGCTG
 ACCCCCCCTGTGCGTGCACCTGCACCAACCTGAAGAACGCCACCAACACCAAGAGC
 AGCAACTGGAAGGAGATGGACCGGGCGAGATCAAGAACTGCAGCTCAAGGTGGCGCC
 GGCAAGCTGATCAACTGCAACACCAGCGTGTACCCAGGCCCTGCCCAAGGTGAGCTTC
 GAGCCCATCCCCATCCACTACTGCGCCCCCGCCGCTTCGCCATCCTGAAGTGCAACGAC
 AAGAAGTTCAACGGCAGGGCCCCCTGCACCAACGTGAGCACCGTGCAGTGCACCCACGGC
 ATCCGGCCCGTGGTGAGCACCCAGCTGCTGTAACGGCAGGCCGGAGGAGGGCGTG
 GTGATCCGAGCGAGAACCTCACCGACAACGCCAACGACATCATCGTGCAGCTGAAGGAG
 AGCGTGGAGATCAACTGCACCCGGCCCCAACAAACACACCCGCAAGAGCATCACCATGGC
 CCCGGCCGCGCCTTACGCCACCGCGACATCATCGCGACATCCGCCAGGCCACTGC
 AACATCAGCGGGCGAGAAGTGGAAACAACACCCCTGAAGCAGATCGTACCGAAGCTGCAGGCC
 CAGTCGGCAACAAGACCATCGTGTCAAGCAGAGCAGCGGGCGACCCCGAGATCGT
 ATGCACAGCTCAACTGCCGGGGAGTTCTTACTGCAACAGCACCCAGCTGTCAAC
 AGCACCTGGAACAACACCATCGGCCCCAACAAACACCAACGGCACCATCACCTGCCCTGC
 CGCATCAAGCAGATCATCACCGCTGGCAGGAGGTGGCAAGGCCATGTACGCCCGG
 ATCCGGGCCAGATCCGCTGCAGCAGCAACATCACCGGCTGCTGTAACCGCGACGGC
 GGCAAGGAGATCAGCAACACCAGAGATCTCCGCCCCGGCGGCGACATGCGCGAC
 AACTGGCGAGCGAGCTGTACAAGTACAAGGTGGTAAGATCGAGCCCTGGCGTGGCC
 CCCACCAAGGCAAGCGCCGCGTGGTGCAGCGCAGAACAGCGCCGCGTACCCCTGGCG
 ATGTTCTGGGCTTCTGGCGCCCGCAGCACCATGGGCCCGCAGCCTGACCCCTG
 ACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTGCAGCAGCAGAACAAACCTGCTGC
 GCCATCGAGGCCAGCAGCACCTGCTGCAGCTGACCGTGTGGGCATCAAGCAGCTGC
 GCCCGCGTGTGGCGCTACCTGAAGGACCAAGCAGCTGCTGGGCATCTGGGC
 TGCAGCGCAAGCTGATCTGCACCAACCGCCGTGCCCTGGAACGCCAGCTGGAGCAACAG
 AGCCTGGACCAGATCTGGAACAAACATGACCTGGATGGAGTGGAGCGCGAGATCGACAAC
 TACACCAACCTGATCTACACCCCTGATCGAGGAGAGCCAGAACACAGCAGGAGAAGAACGAG
 CAGGAGCTGGAGCTGGACAAGTGGCCAGCCTGGAACTGGTCTGACATCAGCAAG
 TGGCTGTGGTACATCAAGATCTTACATGATCGTGGCGGCCCTGGCTGGCCATCGC
 GTGTTCACCGTGTGAGCATCGTAACCGCGTGCAGCAGGCCCTGAGCTTC
 CAGACCCGCTCCCCGCCCGCGGCCCGACGCCCGAGGGCATCGAGGAGGG
 GGCGAGCGCAGCCGACCGCAGCAGGCCCTGGTGCACGCCCTGCTGCCCTGATCTGG
 GACGACCTGCGCAGCCTGCTGCTGAGCTTACCAACGCCCTGCGCAGCCCTGATCCTGATC
 GCCGCCCGCATCGTGGAGCTGCTGGGCCCGCGCTGGGAGGCCCTGAAGTACTGGGC
 AACCTGCTGCAGTACTGGATCCAGGAGCTGAAGAACAGCGCCGTGAGCCTGTTGACGCC
 ATCGCCATCGCCGTGGCCGAGGGCACCGACCGCATCATCGAGGTGGCCAGCGC
 CGGCCCTTCTGCACATCCCCGCCGCATCCGCCAGGGCTTGCAGCGGCCCTGCTGTAA
 CTCGAGCAAGTCTAGAGAATTCCGCCCCCCCCCCCCCTCTCCCTCCCCCCCC
 TAACGTTACTGGCCGAAGCCGCTTGGATAAGGCCGGTGTGCGTTGTCTATGTTATT
 TTCCACCATATTGCCGTCTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGCTTCTT

FIG. 64A

(SEQ ID NO:76)

77 / 131

GACGAGCATTCTAGGGTCTTCCCTCTGCCAAGGAATGCAAGGTCTGTTGAATGT
CGTGAAGGAAGCAGTCCTCTGAAAGCTCTGAAGACAACAAACGTCTGTAGCGACCCT
TTGCAGGCAGCGAACCCCCCACCTGGCGACAGGTGCTCTGCCAAGGGCAGGTGT
ATAAGATAACACCTGCAAAGGCGACAACCCCAGTGCCACGTTGAGTTGGATAGTTGT
GGAAAGAGTCAAATGGCTCTCTCAAGCGTATTCAACAAGGGCTGAAGGATGCCAGAA
GGTACCCATTGTATGGATCTGATCTGGGGCTCGTGCACATGCTTACATGTGTTA
GTCGAGGTTAAAAAAACGTCTAGGCCCGAACCACGGGACGTGGTTTCCTTGAAA
AACACGATAATACCATGGCGCCCGCCAGCGTGTAGCGCGCCAGCTGGACAAGT
GGGAGAAGATCCGCCTGCCCGGGCAAGAAGAAGTACAAGCTGAAGCACATCGTGT
GGGCCAGCGCGAGCTGGAGCGCTCGCCGTGAACCCCGCCTGCTGGAGGACAGCGAGG
GCTGCCGCCAGATCCTGGCCAGCTGCAGCCAGCCTGCAGACCCGAGCGAGGAGCTGC
GCAGCCTGTACAACACCGTGGCACCCCTGTACTGCGTGCACCAGCGATCGACGTCAAGG
ACACCAAGGAGGCCCTGGAGAAGATCGAGGAGGAGCAGAACAGTCCAAGAAGAAGGCC
AGCAGGCCGCCGCCGCCGGCACCGAACAGCAGCCAGGTGAGCCAGAACACTACCCCA
TCGTGCAGAACCTGCAGGCCAGATGGTGCACCAGGCCATCAGCCCCCGCACCCCTGAACG
CCTGGGTGAAGGTGGAGGAGAACGCCCTCAGCCCCGAGGTGATCCCCATGTTAGCG
CCCTGAGCGAGGCCACCCCGAGGACCTGAACACGATGTTGAACACCGTGGGCC
ACCAGGCCGCATGCAGATGCTGAAGGAGACCATCAACGAGGAGGCCAGTGGGACC
GCGTGCACCCCGTCACGCCGGCCCATCGCCCCGGCAGATGCGCAGGCCGGCA
GCGACATGCCGGCACCAACAGCACCTGCAGGAGCAGATGGCTGGATGACCAACAACC
CCCCCATCCCCGTGGCGAGATCTACAAGCGGTGGATCATCCTGGCCAGGGCCCAAGGAGGCC
TGC GGATGTACAGCCCCACCAGCATCCTGGACATCCGCCAGGGCCCAAGGAGGCC
GCGACTACGTGGACCGCTCTACAAGACCTGCGCCTGAGCAGGCCAGGACGTGA
AGAACTGGATGACCGAGACCCCTGCTGGTGAGAACGCCAACCCGACTGCAAGACCATCC
TGAAGGCTCTGGCCCGGCCACCCCTGGAGGAGATGATGACCGCCTGCCAGGGCGTGG
GCCGCCGCCACAAGGCCCGCTGGCCAGGGCGATGAGCCAGGTGACGAACCCGG
CGACCATCATGATGCAGCGGGCAACTCCGCAACCAGCGGAAGACCGTCAAGTGCTCA
ACTGCGGCAAGGAGGCCACCCGCCAGGAACCTGCCGCCAGGCCAGGCAAGAAGGGCTGCT
GGCGCTGCCGCCAGGGCCACCAAGATGAAGGACTGCACCGAGGCCAGGCCACTTCC
TGGGCAAGATCTGGCCAGCTACAAGGCCGCCAGGAACCTCCTGCAGAGGCC
AGCCCACCGCCCCCGAGGGAGAGCTCCGCTGGCGAGGAGAAGACCAACCCAGCC
AGAAGCAGGAGGCCATGACAAGGAGCTGTACCCCTGACCAAGCCTGCGCAGCCTGTC
GCAACGACCCAGCAGCCAGTAAGAATTCAAGACTCGAGCAAGTCTAGA

FIG. 64B
(SEQ ID NO:76)

78 / 131

FIG. 65C

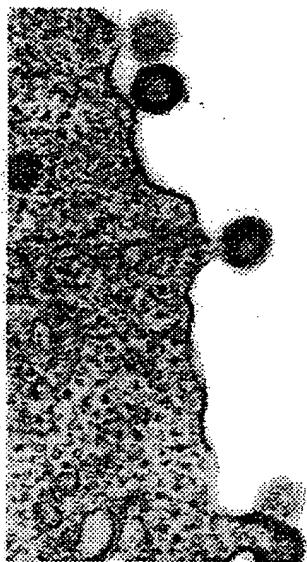


FIG. 65B



FIG. 65A

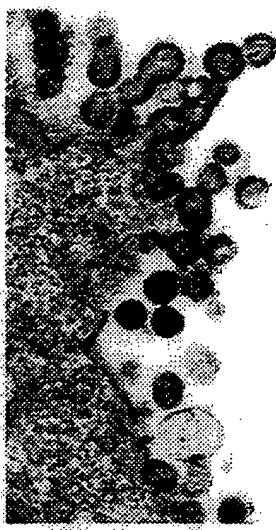


FIG. 65F



FIG. 65E



FIG. 65D

50

gp160 . modSF162	(1)	GAATTCCACCATGGATGCCAATGAAGAGGGCTCTGCTGTGCTGCT
gp160 . modSF162 . delV2	(1)	GAATTCCACCATGGATGCCAATGAAGAGGGCTCTGCTGTGCTGCT
gp160 . modSF162 . delV1V2	(1)	GAATTCCACCATGGATGCCAATGAAGAGGGCTCTGCTGTGCTGCT
gp140 . modSF162	(1)	GAATTCCACCATGGATGCCAATGAAGAGGGCTCTGCTGTGCTGCT
gp140 . mut . modSF162	(1)	GAATTCCACCATGGATGCCAATGAAGAGGGCTCTGCTGTGCTGCT
gp140 . mut7 . modSF162	(1)	GAATTCCACCATGGATGCCAATGAAGAGGGCTCTGCTGTGCTGCT
gp140 . mut8 . modSF162	(1)	GAATTCCACCATGGATGCCAATGAAGAGGGCTCTGCTGTGCTGCT
gp120 . modSF162	(1)	GAATTCCACCATGGATGCCAATGAAGAGGGCTCTGCTGTGCTGCT
Consensus	(1)	GAATTCCACCATGGATGCCAATGAAGAGGGCTCTGCTGTGCTGCT
	51	
gp160 . modSF162	(51)	GCTGTGTGAGCAGTCTTCGTTTCGCCCCAGGCCGCTGGAGAAGCTGTGGGG
gp160 . modSF162 . delV2	(51)	GCTGTGTGAGCAGTCTTCGTTTCGCCCCAGGCCGCTGGAGAAGCTGTGGGG
gp160 . modSF162 . delV1V2	(51)	GCTGTGTGAGCAGTCTTCGTTTCGCCCCAGGCCGCTGGAGAAGCTGTGGGG
gp140 . modSF162	(51)	GCTGTGTGAGCAGTCTTCGTTTCGCCCCAGGCCGCTGGAGAAGCTGTGGGG
gp140 . mut . modSF162	(51)	GCTGTGTGAGCAGTCTTCGTTTCGCCCCAGGCCGCTGGAGAAGCTGTGGGG
gp140 . mut7 . modSF162	(51)	GCTGTGTGAGCAGTCTTCGTTTCGCCCCAGGCCGCTGGAGAAGCTGTGGGG
gp140 . mut8 . modSF162	(51)	GCTGTGTGAGCAGTCTTCGTTTCGCCCCAGGCCGCTGGAGAAGCTGTGGGG
gp120 . modSF162	(51)	GCTGTGTGAGCAGTCTTCGTTTCGCCCCAGGCCGCTGGAGAAGCTGTGGGG
Consensus	(51)	GCTGTGTGAGCAGTCTTCGTTTCGCCCCAGGCCGCTGGAGAAGCTGTGGGG
	101	
gp160 . modSF162	(101)	TGACCGGTGTTACTACGGCGTGCCTGTGGAAGGGGCCACCAACCTG
gp160 . modSF162 . delV2	(101)	TGACCGGTGTTACTACGGCGTGCCTGTGGAAGGGGCCACCAACCTG
gp160 . modSF162 . delV1V2	(101)	TGACCGGTGTTACTACGGCGTGCCTGTGGAAGGGGCCACCAACCTG
gp140 . modSF162	(101)	TGACCGGTGTTACTACGGCGTGCCTGTGGAAGGGGCCACCAACCTG
gp140 . mut . modSF162	(101)	TGACCGGTGTTACTACGGCGTGCCTGTGGAAGGGGCCACCAACCTG
gp140 . mut7 . modSF162	(101)	TGACCGGTGTTACTACGGCGTGCCTGTGGAAGGGGCCACCAACCTG
gp140 . mut8 . modSF162	(101)	TGACCGGTGTTACTACGGCGTGCCTGTGGAAGGGGCCACCAACCTG
gp120 . modSF162	(101)	TGACCGGTGTTACTACGGCGTGCCTGTGGAAGGGGCCACCAACCTG
Consensus	(101)	TGACCGGTGTTACTACGGCGTGCCTGTGGAAGGGGCCACCAACCTG
	150	

FIG. 66A-1

gp160.modSF162	(151)	TTCAGGCGCCAAAGGCCTACGGACACCCGAGGTGCAACAACGTTG	200
gp160.modSF162.delV2	(151)	TTCAGGCGCCAAAGGCCTACGGACACCCGAGGTGCAACAACGTTG	
gp160.modSF162.delV1V2	(151)	TTCAGGCGCCAAAGGCCTACGGACACCCGAGGTGCAACAACGTTG	
gp140.modSF162.delV1V2	(151)	TTCAGGCGCCAAAGGCCTACGGACACCCGAGGTGCAACAACGTTG	
gp140.modSF162	(151)	TTCAGGCGCCAAAGGCCTACGGACACCCGAGGTGCAACAACGTTG	
gp140.mut.modSF162	(151)	TTCAGGCGCCAAAGGCCTACGGACACCCGAGGTGCAACAACGTTG	
gp140.mut7.modSF162	(151)	TTCAGGCGCCAAAGGCCTACGGACACCCGAGGTGCAACAACGTTG	
gp140.mut8.modSF162	(151)	TTCAGGCGCCAAAGGCCTACGGACACCCGAGGTGCAACAACGTTG	
gp120.modSF162	(151)	TTCAGGCGCCAAAGGCCTACGGACACCCGAGGTGCAACAACGTTG	
Consensus	(151)	TTCAGGCGCCAAAGGCCTACGGACACCCGAGGTGCAACAACGTTG	
	201		
gp160.modSF162	(201)	GGCCACCCACGGCTTGGTCCCCACGGAGCCCCAACGGAGATGGTG	
gp160.modSF162.delV2	(201)	GGCCACCCACGGCTTGGTCCCCACGGAGCCCCAACGGAGATGGTG	
gp160.modSF162.delV1V2	(201)	GGCCACCCACGGCTTGGTCCCCACGGAGCCCCAACGGAGATGGTG	
gp140.modSF162	(201)	GGCCACCCACGGCTTGGTCCCCACGGAGCCCCAACGGAGATGGTG	
gp140.mut.modSF162	(201)	GGCCACCCACGGCTTGGTCCCCACGGAGCCCCAACGGAGATGGTG	
gp140.mut7.modSF162	(201)	GGCCACCCACGGCTTGGTCCCCACGGAGCCCCAACGGAGATGGTG	
gp140.mut8.modSF162	(201)	GGCCACCCACGGCTTGGTCCCCACGGAGCCCCAACGGAGATGGTG	
gp120.modSF162	(201)	GGCCACCCACGGCTTGGTCCCCACGGAGCCCCAACGGAGATGGTG	
Consensus	(201)	GGCCACCCACGGCTTGGTCCCCACGGAGCCCCAACGGAGATGGTG	
	251		
gp160.modSF162	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACACATGGTGGAG	
gp160.modSF162.delV2	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACACATGGTGGAG	
gp160.modSF162.delV1V2	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACACATGGTGGAG	
gp140.modSF162	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACACATGGTGGAG	
gp140.mut.modSF162	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACACATGGTGGAG	
gp140.mut7.modSF162	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACACATGGTGGAG	
gp140.mut8.modSF162	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACACATGGTGGAG	

FIG. 66A-2

gp120 .modSF162 Consensus	(251) TGGAGAACGTGACCGAGAACTCAACATGTGGAAAACAACATGGTGGAG (251) TGGAGAACGTGACCGAGAACTCAACATGTGGAAAACAACATGGTGGAG 301
gp160 .modSF162 gp160 .modSF162 .delV2 gp160 .modSF162 .delV1V2 gp140 .modSF162 gp140 .mut .modSF162 gp140 .mut7 .modSF162 gp140 .mut8 .modSF162 gp120 .modSF162 Consensus	(301) CAGATGCACCGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTG (301) CAGATGCACCGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTG (301) CAGATGCACCGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTG (301) CAGATGCACCGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTG (301) CAGATGCACCGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTG (301) CAGATGCACCGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTG (301) CAGATGCACCGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTG 350
gp160 .modSF162 gp160 .modSF162 .delV2 gp160 .modSF162 .delV1V2 gp140 .modSF162 gp140 .mut .modSF162 gp140 .mut7 .modSF162 gp140 .mut8 .modSF162 gp120 .modSF162 Consensus	(351) CGTGAAGCTGACCCCCCTGTGGTGAACCCCTGCACCTGCACCAACCTGAAGA (351) CGTGAAGCTGACCCCCCTGTGGTGAACCCCTGCACCTGCACCAACCTGAAGA (351) CGTGAAGCTGACCCCCCTGTGGTGAACCCCTGCACCTGCACCAACCTGAAGA (351) CGTGAAGCTGACCCCCCTGTGGTGAACCCCTGCACCTGCACCAACCTGAAGA (351) CGTGAAGCTGACCCCCCTGTGGTGAACCCCTGCACCTGCACCAACCTGAAGA (351) CGTGAAGCTGACCCCCCTGTGGTGAACCCCTGCACCTGCACCAACCTGAAGA (351) CGTGAAGCTGACCCCCCTGTGGTGAACCCCTGCACCTGCACCAACCTGAAGA 400
gp160 .modSF162 gp160 .modSF162 .delV2 gp160 .modSF162 .delV1V2 gp140 .modSF162 gp140 .mut .modSF162 gp140 .mut7 .modSF162 gp140 .mut8 .modSF162 gp120 .modSF162 Consensus	(351) ACGCCACCAAACACCAAGAGCAGCAACTGGAAAGGAGATGGACCCGGCGAG (401) ACGCCACCAAACACCAAGAGCAGCAACTGGAAAGGAGATGGACCCGGCGAG (375) ----- (401) ACGCCACCAAACACCAAGAGCAGCAACTGGAAAGGAGATGGACCCGGCGAG (401) ACGCCACCAAACACCAAGAGCAGCAACTGGAAAGGAGATGGACCCGGCGAG (401) ACGCCACCAAACACCAAGAGCAGCAACTGGAAAGGAGATGGACCCGGCGAG (401) ACGCCACCAAACACCAAGAGCAGCAACTGGAAAGGAGATGGACCCGGCGAG 450
gp160 .modSF162 gp160 .modSF162 .delV2 gp160 .modSF162 .delV1V2 gp140 .modSF162 gp140 .mut .modSF162 gp140 .mut7 .modSF162 gp140 .mut8 .modSF162 gp120 .modSF162 Consensus	(401) ACGCCACCAAACACCAAGAGCAGCAACTGGAAAGGAGATGGACCCGGCGAG (401) ACGCCACCAAACACCAAGAGCAGCAACTGGAAAGGAGATGGACCCGGCGAG (401) ACGCCACCAAACACCAAGAGCAGCAACTGGAAAGGAGATGGACCCGGCGAG (401) ACGCCACCAAACACCAAGAGCAGCAACTGGAAAGGAGATGGACCCGGCGAG (401) ACGCCACCAAACACCAAGAGCAGCAACTGGAAAGGAGATGGACCCGGCGAG 500

FIG. 66A-3

82 / 131

FIG. 66A-4

gp140 .mut7 .modSF162	(551)	ACAAACACCAGCTACAAGCTGATCAACTGCAAACACCAGGGTGTACCCCCAG	
gp140 .mut8 .modSF162	(551)	ACAAACACCAGCTACAAGCTGATCAACTGCAAACACCAGGGTGTACCCCCAG	
gp120 .modSF162	(551)	ACAAACACCAGCTACAAGCTGATCAACTGCAAACACCAGGGTGTACCCCCAG	650
Consensus	601	ACAAACACCAGCTACAAGCTGATCAACTGCAAACACCAGGGTGTACCCCCAG	
gp160 .modSF162	(601)	GCCTGCCCCAAGGGTAGGCTTCGAGCCCCATCCCACTACTGGCCCC	
gp160 .modSF162 .delV2	(520)	GCCTGCCCCAAGGGTAGGCTTCGAGCCCCATCCCACTACTGGCCCC	
gp160 .modSF162 .delV1V2	(412)	GCCTGCCCCAAGGGTAGGCTTCGAGCCCCATCCCACTACTGGCCCC	
gp140 .modSF162	(601)	GCCTGCCCCAAGGGTAGGCTTCGAGCCCCATCCCACTACTGGCCCC	
gp140 .mut .modSF162	(601)	GCCTGCCCCAAGGGTAGGCTTCGAGCCCCATCCCACTACTGGCCCC	
gp140 .mut7 .modSF162	(601)	GCCTGCCCCAAGGGTAGGCTTCGAGCCCCATCCCACTACTGGCCCC	
gp140 .mut8 .modSF162	(601)	GCCTGCCCCAAGGGTAGGCTTCGAGCCCCATCCCACTACTGGCCCC	
gp120 .modSF162	(601)	GCCTGCCCCAAGGGTAGGCTTCGAGCCCCATCCCACTACTGGCCCC	
Consensus	651	GCCTGCCCCAAGGGTAGGCTTCGAGCCCCATCCCACTACTGGCCCC	700
gp160 .modSF162	(651)	CGCCGGCTTCGCCATCCTGAAGTGC AACGACAAGTCAACGGCAGCG	
gp160 .modSF162 .delV2	(570)	CGCCGGCTTCGCCATCCTGAAGTGC AACGACAAGTCAACGGCAGCG	
gp160 .modSF162 .delV1V2	(462)	CGCCGGCTTCGCCATCCTGAAGTGC AACGACAAGTCAACGGCAGCG	
gp140 .modSF162	(651)	CGCCGGCTTCGCCATCCTGAAGTGC AACGACAAGTCAACGGCAGCG	
gp140 .mut .modSF162	(651)	CGCCGGCTTCGCCATCCTGAAGTGC AACGACAAGTCAACGGCAGCG	
gp140 .mut7 .modSF162	(651)	CGCCGGCTTCGCCATCCTGAAGTGC AACGACAAGTCAACGGCAGCG	
gp140 .mut8 .modSF162	(651)	CGCCGGCTTCGCCATCCTGAAGTGC AACGACAAGTCAACGGCAGCG	
gp120 .modSF162	(651)	CGCCGGCTTCGCCATCCTGAAGTGC AACGACAAGTCAACGGCAGCG	
Consensus	701	CGCCGGCTTCGCCATCCTGAAGTGC AACGACAAGTCAACGGCAGCG	750
gp160 .modSF162	(701)	GCCCCCTGCACCAACGGTGCAGTGCACCCACGGCATCCGGCCCC	
gp160 .modSF162 .delV2	(620)	GCCCCCTGCACCAACGGTGCAGTGCACCCACGGCATCCGGCCCC	
gp160 .modSF162 .delV1V2	(512)	GCCCCCTGCACCAACGGTGCAGTGCACCCACGGCATCCGGCCCC	
gp140 .modSF162	(701)	GCCCCCTGCACCAACGGTGCAGTGCACCCACGGCATCCGGCCCC	
gp140 .mut .modSF162	(701)	GCCCCCTGCACCAACGGTGCAGTGCACCCACGGCATCCGGCCCC	
gp140 .mut7 .modSF162	(701)	GCCCCCTGCACCAACGGTGCAGTGCACCCACGGCATCCGGCCCC	
gp140 .mut8 .modSF162	(701)	GCCCCCTGCACCAACGGTGCAGTGCACCCACGGCATCCGGCCCC	
gp120 .modSF162	(701)	GCCCCCTGCACCAACGGTGCAGTGCACCCACGGCATCCGGCCCC	
Consensus	(701)	GCCCCCTGCACCAACGGTGCAGTGCACCCACGGCATCCGGCCCC	

FIG. 66A-5

		800
gp160.modSF162	(751)	GTTGGTGAGGCCACCCAGCTGCTGAACGGCAGCCTGGCGAGGGGGGT
gp160.modSF162.delV2	(670)	GTTGGTGGCACCCAGCTGGCTGAACGGCAGCCTGGCGAGGGGGGT
gp160.modSF162.delV1V2	(562)	GTTGGTGGCACCCAGCTGGCTGAACGGCAGCCTGGCGAGGGGGGT
gp160.modSF162	(751)	GTTGGTGGCACCCAGCTGGCTGAACGGCAGCCTGGCGAGGGGGGT
gp140.modSF162	(751)	GTTGGTGGCACCCAGCTGGCTGAACGGCAGCCTGGCGAGGGGGGT
gp140.mut.modSF162	(751)	GTTGGTGGCACCCAGCTGGCTGAACGGCAGCCTGGCGAGGGGGGT
gp140.mut7.modSF162	(751)	GTTGGTGGCACCCAGCTGGCTGAACGGCAGCCTGGCGAGGGGGGT
gp140.mut8.modSF162	(751)	GTTGGTGGCACCCAGCTGGCTGAACGGCAGCCTGGCGAGGGGGGT
gp120.modSF162	(751)	GTTGGTGGCACCCAGCTGGCTGAACGGCAGCCTGGCGAGGGGGGT
Consensus	(751)	GTTGGTGGCACCCAGCTGGCTGAACGGCAGCCTGGCGAGGGGGGT
	801	
gp160.modSF162	(801)	GGTGATCCGCAGCGAGAAACTTACCGGACAAGGCCAACCATCATCGTGC
gp160.modSF162.delV2	(720)	GGTGATCCGCAGCGAGAACTTACCGGACAAGGCCAACCATCATCGTGC
gp160.modSF162.delV1V2	(612)	GGTGATCCGCAGCGAGAACTTACCGGACAAGGCCAACCATCATCGTGC
gp140.modSF162	(801)	GGTGATCCGCAGCGAGAACTTACCGGACAAGGCCAACCATCATCGTGC
gp140.mut.modSF162	(801)	GGTGATCCGCAGCGAGAACTTACCGGACAAGGCCAACCATCATCGTGC
gp140.mut7.modSF162	(801)	GGTGATCCGCAGCGAGAACTTACCGGACAAGGCCAACCATCATCGTGC
gp140.mut8.modSF162	(801)	GGTGATCCGCAGCGAGAACTTACCGGACAAGGCCAACCATCATCGTGC
gp120.modSF162	(801)	GGTGATCCGCAGCGAGAACTTACCGGACAAGGCCAACCATCATCGTGC
Consensus	(801)	GGTGATCCGCAGCGAGAACTTACCGGACAAGGCCAACCATCATCGTGC
	851	
gp160.modSE162	(851)	AGCTGAAGGGAGGGTGGAGATCAACTGACCCGGCCCCAACAAACACC
gp160.modSF162.delV2	(770)	AGCTGAAGGGAGGGTGGAGATCAACTGACCCGGCCCCAACAAACACC
gp160.modSF162.delV1V2	(662)	AGCTGAAGGGAGGGTGGAGATCAACTGACCCGGCCCCAACAAACACC

FIG. 66A-6

gp140.modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGCACCCGCCAACAAACACC
gp140.mut.modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGCACCCGCCAACAAACACC
gp140.mut7.modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGCACCCGCCAACAAACACC
gp140.mut8.modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGCACCCGCCAACAAACACC
gp120.modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGCACCCGCCAACAAACACC
Consensus	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGCACCCGCCAACAAACACC
	950	
gp160.modSF162	(901)	CGCAAGAGCATCACCATCGCCCCGGCGCCTCTACGCCACCGGGCA
gp160.modSF162.delV2	(820)	CGCAAGAGCATCACCATCGCCCCGGCGCCTCTACGCCACCGGGCA
gp160.modSF162.delV1V2	(712)	CGCAAGAGCATCACCATCGCCCCGGCGCCTCTACGCCACCGGGCA
gp140.modSF162	(901)	CGCAAGAGCATCACCATCGCCCCGGCGCCTCTACGCCACCGGGCA
gp140.mut.modSF162	(901)	CGCAAGAGCATCACCATCGCCCCGGCGCCTCTACGCCACCGGGCA
gp140.mut7.modSF162	(901)	CGCAAGAGCATCACCATCGCCCCGGCGCCTCTACGCCACCGGGCA
gp140.mut8.modSF162	(901)	CGCAAGAGCATCACCATCGCCCCGGCGCCTCTACGCCACCGGGCA
gp140.mut8.modSF162	(901)	CGCAAGAGCATCACCATCGCCCCGGCGCCTCTACGCCACCGGGCA
gp120.modSF162	(901)	CGCAAGAGCATCACCATCGCCCCGGCGCCTCTACGCCACCGGGCA
Consensus	(901)	CGCAAGAGCATCACCATCGCCCCGGCGCCTCTACGCCACCGGGCA
	951	
gp160.modSF162	(951)	CATCATCGGGGACATCGCCAAACATCGCAACATCGGGGGAGAAGT
gp160.modSF162.delV2	(870)	CATCATCGGGGACATCGCCAAACATCGCAACATCGGGGGAGAAGT
gp160.modSF162.delV1V2	(762)	CATCATCGGGGACATCGCCAAACATCGCAACATCGGGGGAGAAGT
gp140.modSF162	(951)	CATCATCGGGGACATCGCCAAACATCGCAACATCGGGGGAGAAGT
gp140.mut.modSF162	(951)	CATCATCGGGGACATCGCCAAACATCGCAACATCGGGGGAGAAGT
gp140.mut7.modSF162	(951)	CATCATCGGGGACATCGCCAAACATCGCAACATCGGGGGAGAAGT
gp140.mut8.modSF162	(951)	CATCATCGGGGACATCGCCAAACATCGCAACATCGGGGGAGAAGT
gp120.modSF162	(951)	CATCATCGGGGACATCGCCAAACATCGCAACATCGGGGGAGAAGT
Consensus	(951)	CATCATCGGGGACATCGCCAAACATCGCAACATCGGGGGAGAAGT
	1000	

FIG. 66A-7

				1050
gp160.modSF162	(1001)	GGAAACAAACACCCCTGAAGGCAGATCGTGACCAAGGCCAGGTGCAAGGCCAGTTCGGC		
gp160.modSF162.delV2	(920)	GGAAACAAACACCCCTGAAGGCAGATCGTGACCAAGGCCAGTTCGGC		
gp160.modSF162.delV1V2	(812)	GGAAACAAACACCCCTGAAGGCAGATCGTGACCAAGGCCAGTTCGGC		
gp140.modSF162	(1001)	GGAAACAAACACCCCTGAAGGCAGATCGTGACCAAGGCCAGTTCGGC		
gp140.mut.modSF162	(1001)	GGAAACAAACACCCCTGAAGGCAGATCGTGACCAAGGCCAGTTCGGC		
gp140.mut7.modSF162	(1001)	GGAAACAAACACCCCTGAAGGCAGATCGTGACCAAGGCCAGTTCGGC		
gp140.mut8.modSF162	(1001)	GGAAACAAACACCCCTGAAGGCAGATCGTGACCAAGGCCAGTTCGGC		
gp120.modsSF162	(1001)	GGAAACAAACACCCCTGAAGGCAGATCGTGACCAAGCTGAGGGCCAGTTCGGC		
Consensus		GGAAACAAACACCCCTGAAGGCAGATCGTGACCAAGCTGAGGGCCAGTTCGGC		1100
			1051	
gp160.modSF162	(1051)	AACAAGACCATACTGTGTTCAAGCAGAGCCAGGGGGGAGCCCGAGATCGT		
gp160.modSF162.delV2	(970)	AACAAGACCATACTGTGTTCAAGCAGAGCCAGGGGGGAGCCCGAGATCGT		
gp160.modSF162.delV1V2	(862)	AACAAGACCATACTGTGTTCAAGCAGAGCCAGGGGGGAGCCCGAGATCGT		
gp140.modSF162	(1051)	AACAAGACCATACTGTGTTCAAGCAGAGCCAGGGGGGAGCCCGAGATCGT		
gp140.mut.modsSF162	(1051)	AACAAGACCATACTGTGTTCAAGCAGAGCCAGGGGGGAGCCCGAGATCGT		
gp140.mut7.modSF162	(1051)	AACAAGACCATACTGTGTTCAAGCAGAGCCAGGGGGGAGCCCGAGATCGT		
gp140.mut8.modSF162	(1051)	AACAAGACCATACTGTGTTCAAGCAGAGCCAGGGGGGAGCCCGAGATCGT		
gp120.modsSF162	(1051)	AACAAGACCATACTGTGTTCAAGCAGAGCCAGGGGGGAGCCCGAGATCGT		
Consensus		AACAAGACCATACTGTGTTCAAGCAGAGCCAGGGGGGAGCCCGAGATCGT		1150
			1101	
gp160.modSF162	(1101)	GATGCACAGCTTCAACTGGGGGAGTTCTTACTTGCAACAGCACCC		
gp160.modSF162.delV2	(1020)	GATGCACAGCTTCAACTGGGGGAGTTCTTACTTGCAACAGCACCC		
gp160.modSF162.delV1V2	(912)	GATGCACAGCTTCAACTGGGGGAGTTCTTACTTGCAACAGCACCC		
gp140.modSF162	(1101)	GATGCACAGCTTCAACTGGGGGAGTTCTTACTTGCAACAGCACCC		
gp140.mut.modsSF162	(1101)	GATGCACAGCTTCAACTGGGGGAGTTCTTACTTGCAACAGCACCC		
gp140.mut7.modsSF162	(1101)	GATGCACAGCTTCAACTGGGGGAGTTCTTACTTGCAACAGCACCC		
gp140.mut8.modsSF162	(1101)	GATGCACAGCTTCAACTGGGGGAGTTCTTACTTGCAACAGCACCC		
gp120.modsSF162	(1101)	GATGCACAGCTTCAACTGGGGGAGTTCTTACTTGCAACAGCACCC		
Consensus		GATGCACAGCTTCAACTGGGGGAGTTCTTACTTGCAACAGCACCC		1200
gp160.modsSF162	(1151)	AGCTGTTCAACAGCACCTGGAACAAACACCCATGGCCCCAACACCAAAC		
			1151	

FIG. 66A-8

gp160.modSF162.delV2	(1070)	AGCTGTTCAACAGCACCTGGAAACAACACATGGCCCCAACAAACCCAC
gp160.modSF162.delV2	(962)	AGCTGTTCAACAGCACCTGGAAACAACACATGGCCCCAACAAACCCAC
gp140.modSF162	(1151)	AGCTGTTCAACAGCACCTGGAAACAACACATGGCCCCAACAAACCCAC
gp140.mut.modSF162	(1151)	AGCTGTTCAACAGCACCTGGAAACAACACATGGCCCCAACAAACCCAC
gp140.mut.modSF162	(1151)	AGCTGTTCAACAGCACCTGGAAACAACACATGGCCCCAACAAACCCAC
gp140.mut7.modSF162	(1151)	AGCTGTTCAACAGCACCTGGAAACAACACATGGCCCCAACAAACCCAC
gp140.mut8.modSF162	(1151)	AGCTGTTCAACAGCACCTGGAAACAACACATGGCCCCAACAAACCCAC
gp120.modSF162	(1151)	AGCTGTTCAACAGCACCTGGAAACAACACATGGCCCCAACAAACCCAC
Consensus	(1151)	AGCTGTTCAACAGCACCTGGAAACAACACATGGCCCCAACAAACCCAC
	1201	1250
gp160.modSF162	(1201)	GCGACCATCACCCGTGCCATCAAGCAGATCATCAACCGCTGGCA
gp160.modSF162.delV2	(1120)	GCGACCATCACCCGTGCCATCAAGCAGATCATCAACCGCTGGCA
gp160.modSF162.delV2	(1012)	GCGACCATCACCCGTGCCCTGCGGCATCAAGCAGATCATCAACCGCTGGCA
gp140.modSF162	(1201)	GCGACCATCACCCGTGCCCTGCGGCATCAAGCAGATCATCAACCGCTGGCA
gp140.mut.modSF162	(1201)	GCGACCATCACCCGTGCCCTGCGGCATCAAGCAGATCATCAACCGCTGGCA
gp140.mut7.modSF162	(1201)	GCGACCATCACCCGTGCCCTGCGGCATCAAGCAGATCATCAACCGCTGGCA
gp140.mut8.modSF162	(1201)	GCGACCATCACCCGTGCCCTGCGGCATCAAGCAGATCATCAACCGCTGGCA
gp120.modSF162	(1201)	GCGACCATCACCCGTGCCCTGCGGCATCAAGCAGATCATCAACCGCTGGCA
Consensus	(1201)	GCGACCATCACCCGTGCCCTGCGGCATCAAGCAGATCATCAACCGCTGGCA
	1251	1300
gp160.modSF162	(1251)	GGAGGTTGGCAAGGCCATGTACGCCCATCGGGCCAGATCCG
gp160.modSF162.delV2	(1170)	GGAGGTTGGCAAGGCCATGTACGCCCATCGGGCCAGATCCG
gp160.modSF162.delV2	(1062)	GGAGGTTGGCAAGGCCATGTACGCCCATCGGGCCAGATCCG
gp140.modSF162	(1251)	GGAGGTTGGCAAGGCCATGTACGCCCATCGGGCCAGATCCG
gp140.mut.modSF162	(1251)	GGAGGTTGGCAAGGCCATGTACGCCCATCGGGCCAGATCCG
gp140.mut7.modSF162	(1251)	GGAGGTTGGCAAGGCCATGTACGCCCATCGGGCCAGATCCG
gp140.mut8.modSF162	(1251)	GGAGGTTGGCAAGGCCATGTACGCCCATCGGGCCAGATCCG
gp120.modSF162	(1251)	GGAGGTTGGCAAGGCCATGTACGCCCATCGGGCCAGATCCG
Consensus	(1251)	GGAGGTTGGCAAGGCCATGTACGCCCATCGGGCCAGATCCG

FIG. 66A-9

FIG. 66A-10

gp160 .modSF162	(1451)	TGGCCGTGGCCCCCAAGGCCAAGGCCAGGCCGGTGGTGCAGGCCGAGAAG	1500
gp160 .modSF162 .delV2	(1370)	TGGCGTGGCCCCCAAGCCAAGGCCAAGGCCAGGCCGGTGGTGCAGGCCGAGAAG	
gp160 .modSF162 .delV1V2	(1262)	TGGCGTGGCCCCCAAGGCCAAGGCCAAGGCCAGGCCGGTGGTGCAGGCCGAGAAG	
gp140 .modSF162	(1451)	TGGCGTGGCCCCCAAGGCCAAGGCCAAGGCCAGGCCGGTGGTGCAGGCCGAGAAG	
gp140 .mut .modSF162	(1451)	TGGCGTGGCCCCCAAGGCCAAGGCCAAGGCCAGGCCGGTGGTGCAGGCCGAGAAG	
gp140 .mut7 .modSF162	(1451)	TGGCGTGGCCCCCAAGGCCAAGGCCATAGCAGCGTGGTGCAGGCCGAGAAG	
gp140 .mut8 .modSF162	(1451)	TGGCGTGGCCCCCACCATGCCATAGCAGCGTGGTGCAGGCCGAGAAG	
gp120 .modSF162	(1451)	TGGCGTGGCCCCACCAAGGCCAAGGCCAAGGCCAGGCCGGTGGTGCAGGCCGAGAAG	
Consensus	(1451)	TGGCGTGGCCCCACCAAGGCCAAGGCCAAGGCCAGGCCGGTGGTGCAGGCCGAGAAG	
	1501		1500
gp160 .modSF162	(1501)	CGGCCGTGACCCCTGGGGGCCATGTTCTGGGCTTCTGGGCTTCTGGGCGCCGG	
gp160 .modSF162 .delV2	(1420)	CGGCCGTGACCCCTGGGGGCCATGTTCTGGGCTTCTGGGCTTCTGGGCGCCGG	
gp160 .modSF162 .delV1V2	(1312)	CGGCCGTGACCCCTGGGGGCCATGTTCTGGGCTTCTGGGCTTCTGGGCGCCGG	
gp140 .modSF162	(1501)	CGGCCGTGACCCCTGGGGGCCATGTTCTGGGCTTCTGGGCGCCGG	
gp140 .mut .modSF162	(1501)	AGGCCGTGACCCCTGGGGGCCATGTTCTGGGCTTCTGGGCGCCGG	
gp140 .mut7 .modSF162	(1501)	AGGCCGTGACCCCTGGGGGCCATGTTCTGGGCTTCTGGGCGCCGG	
gp140 .mut8 .modSF162	(1501)	AGGCCGTGACCCCTGGGGGCCATGTTCTGGGCTTCTGGGCGCCGG	
gp120 .modSF162	(1501)	CGC---TAACTCGA-----	
Consensus	(1501)	CGGCCGTGACCCCTGGGGGCCATGTTCTGGGCTTCTGGGCGCCGG	
	1551		1600
gp160 .modSF162	(1551)	CAGCACCATGGGCCGCCAGCCCTGACCCCTGACCGTGCAGGCCGCCAGC	
gp160 .modSF162 .delV2	(1470)	CAGCACCATGGGCCGCCAGCCCTGACCCCTGACCGTGCAGGCCGCCAGC	
gp160 .modSF162 .delV1V2	(1362)	CAGCACCATGGGCCGCCAGCCCTGACCCCTGACCGTGCAGGCCGCCAGC	
gp140 .modSF162	(1551)	CAGCACCATGGGCCGCCAGCCCTGACCCCTGACCGTGCAGGCCGCCAGC	
gp140 .mut .modSF162	(1551)	CAGCACCATGGGCCGCCAGCCCTGACCCCTGACCGTGCAGGCCGCCAGC	
gp140 .mut7 .modSF162	(1551)	CAGCACCATGGGCCGCCAGCCCTGACCCCTGACCGTGCAGGCCGCCAGC	
gp140 .mut8 .modSF162	(1551)	CAGCACCATGGGCCGCCAGCCCTGACCCCTGACCGTGCAGGCCGCCAGC	
gp120 .modSF162	(1513)	-----	
Consensus	(1551)	CAGCACCATGGGCCGCCAGCCCTGACCCCTGACCGTGCAGGCCGCCAGC	

FIG. 66A-11

gp160 .modSF162 gp160 .modSF162 .delV2 gp160 .modSF162 .delV2 .delV2 gp140 .modSF162 gp140 .mut .modSF162 gp140 .mut7 .modSF162 gp140 .mut8 .modSF162 gp120 .modSF162 Consensus	1601 (1601) TGCTGAGCGGCATCGTGCAGCAGAACAAACCTGCTGGCGGCCATCGAG (1520) TGCTGAGCGGCATCGTGCAGCAGAACAAACCTGCTGGCGGCCATCGAG (1412) TGCTGAGCGGCATCGTGCAGCAGAACAAACCTGCTGGCGGCCATCGAG (1601) TGCTGAGCGGCATCGTGCAGCAGAACAAACCTGCTGGCGGCCATCGAG (1601) TGCTGAGCGGCATCGTGCAGCAGAACAAACCTGCTGGCGGCCATCGAG (1601) TGCTGAGCGGCATCGTGCAGCAGAACAAACCTGCTGGCGGCCATCGAG (1601) TGCTGAGCGGCATCGTGCAGCAGAACAAACCTGCTGGCGGCCATCGAG (1513) ----- (1601) TGCTGAGCGGCATCGTGCAGCAGAACAAACCTGCTGGCGGCCATCGAG	1650 1651	1651 (1651) GCCCAGGAGCACCTGCTGCAGCTGACCGTTGGGGCATCAAGCAGCTGCA (1570) GCCCAGGAGCACCTGCTGCAGCTGACCGTTGGGGCATCAAGCAGCTGCA (1462) GCCCAGGAGCACCTGCTGCAGCTGACCGTTGGGGCATCAAGCAGCTGCA (1651) GCCCAGGAGCACCTGCTGCAGCTGACCGTTGGGGCATCAAGCAGCTGCA (1651) GCCCAGGAGCACCTGCTGCAGCTGACCGTTGGGGCATCAAGCAGCTGCA (1651) GCCCAGGAGCACCTGCTGCAGCTGACCGTTGGGGCATCAAGCAGCTGCA (1651) GCCCAGGAGCACCTGCTGCAGCTGACCGTTGGGGCATCAAGCAGCTGCA (1513) ----- (1651) GCCCAGGAGCACCTGCTGCAGCTGACCGTTGGGGCATCAAGCAGCTGCA	1700 1701	1701 (1701) GGCCCCGGCTGTGGCCGCTGGAGGGCTTACCTGAAGGGACCAGCAGCTGCTGG (1620) GGCCCCGGCTGTGGCCGCTGGAGGGCTTACCTGAAGGGACCAGCAGCTGCTGG (1512) GGCCCCGGCTGTGGCCGCTGGAGGGCTTACCTGAAGGGACCAGCAGCTGCTGG (1701) GGCCCCGGCTGTGGCCGCTGGAGGGCTTACCTGAAGGGACCAGCAGCTGCTGG (1701) GGCCCCGGCTGTGGCCGCTGGAGGGCTTACCTGAAGGGACCAGCAGCTGCTGG (1701) GGCCCCGGCTGTGGCCGCTGGAGGGCTTACCTGAAGGGACCAGCAGCTGCTGG
---	--	--------------	--	--------------	--

FIG. 66A-12

91 / 131

gp120.modSF162	(1513)	-----	GGCCCGCGTGTGGCGCTACCTGAAGGACCAGCAGCTGCTGG
Consensus	(1701)	1751	
gp160.modSF162	(1751)	GCATCTGGGGCTGCAGGGCAAGCTGATCTGCACCCGGCTGCCCTGG	
gp160.modSF162.de1V2	(1670)	GCATCTGGGGCTGCAGGGCAAGCTGATCTGCACCCGGCTGCCCTGG	
gp160.modSF162.de1V1V2	(1562)	GCATCTGGGGCTGCAGGGCAAGCTGATCTGCACCCGGCTGCCCTGG	
gp140.modSF162	(1751)	GCATCTGGGGCTGCAGGGCAAGCTGATCTGCACCCGGCTGCCCTGG	
gp140.mut.modSF162	(1751)	GCATCTGGGGCTGCAGGGCAAGCTGATCTGCACCCGGCTGCCCTGG	
gp140.mut7.modSF162	(1751)	GCATCTGGGGCTGCAGGGCAAGCTGATCTGCACCCGGCTGCCCTGG	
gp140.mut8.modSF162	(1751)	GCATCTGGGGCTGCAGGGCAAGCTGATCTGCACCCGGCTGCCCTGG	
gp120.modSF162	(1513)	-----	
Consensus	(1751)	GCATCTGGGGCTGCAGGGCAAGCTGATCTGCACCCGGCTGCCCTGG	
	1801		
gp160.modSF162	(1801)	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGCTGGAAACATGAC	
gp160.modSF162.de1V2	(1720)	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGCTGGAAACATGAC	
gp160.modSF162.de1V1V2	(1612)	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGCTGGAAACATGAC	
gp140.modSF162	(1801)	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGCTGGAAACATGAC	
gp140.mut.modSF162	(1801)	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGCTGGAAACATGAC	
gp140.mut7.modSF162	(1801)	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGCTGGAAACATGAC	
gp140.mut8.modSF162	(1801)	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGCTGGAAACATGAC	
gp120.modSF162	(1513)	-----	
Consensus	(1801)	AACGCCAGCTGGAGCAACAAGAGCCTGGACCAAGCTGGAAACATGAC	
	1851		
gp160.modSF162	(1851)	CTGGATGGAGTGGAGCTGGAGATCGACAACTACACCAACCTGATCTACA	
gp160.modSF162.de1V2	(1770)	CTGGATGGAGTGGAGCTGGAGATCGACAACTACACCAACCTGATCTACA	
gp160.modSF162.de1V1V2	(1662)	CTGGATGGAGTGGAGCTGGAGATCGACAACTACACCAACCTGATCTACA	
gp140.modSF162	(1851)	CTGGATGGAGTGGAGCTGGAGATCGACAACTACACCAACCTGATCTACA	
gp140.mut.modSF162	(1851)	CTGGATGGAGTGGAGCTGGAGATCGACAACTACACCAACCTGATCTACA	
gp140.mut7.modSF162	(1851)	CTGGATGGAGTGGAGCTGGAGATCGACAACTACACCAACCTGATCTACA	
gp140.mut8.modSF162	(1851)	CTGGATGGAGTGGAGCTGGAGATCGACAACTACACCAACCTGATCTACA	
gp120.modSF162	(1513)	-----	
Consensus	(1851)	CTGGATGGAGTGGAGCTGGAGATCGACAACTACACCAACCTGATCTACA	
	1900		

FIG. 66A-13

92 / 131

			1950
gp160.modSF162	(1901)	CCCTGATCGAGGAGGCCAGAACCCAGCAGGAAGAACGGAGCAGGAGCTG	
gp160.modSF162.de1V2	(1820)	CCCTGATCGAGGAGGCCAGAACCCAGCAGGAAGAACGGAGCAGGAGCTG	
gp160.modSF162.de1V2	(1712)	CCCTGATCGAGGAGGCCAGAACCCAGCAGGAAGAACGGAGCAGGAGCTG	
gp140.modSF162	(1901)	CCCTGATCGAGGAGGCCAGAACCCAGCAGGAAGAACGGAGCAGGAGCTG	
gp140.mut.modSF162	(1901)	CCCTGATCGAGGAGGCCAGAACCCAGCAGGAAGAACGGAGCAGGAGCTG	
gp140.mut7.modSF162	(1901)	CCCTGATCGAGGAGGCCAGAACCCAGCAGGAAGAACGGAGCAGGAGCTG	
gp140.mut7.modSF162	(1901)	CCCTGATCGAGGAGGCCAGAACCCAGCAGGAAGAACGGAGCAGGAGCTG	
gp140.mut8.modSF162	(1901)	CCCTGATCGAGGAGGCCAGAACCCAGCAGGAAGAACGGAGCAGGAGCTG	
gp120.modSF162	(1513)	-----	
Consensus	(1901)	CCCTGATCGAGGAGGCCAGAACCCAGCAGGAAGAACGGAGCAGGAGCTG	2000
			1951
gp160.modSF162	(1951)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACTGGTTCGACATCAGCAA	
gp160.modSF162.de1V2	(1870)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACTGGTTCGACATCAGCAA	
gp160.modSF162.de1V2	(1762)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACTGGTTCGACATCAGCAA	
gp140.modSF162	(1951)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACTGGTTCGACATCAGCAA	
gp140.mut.modSF162	(1951)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACTGGTTCGACATCAGCAA	
gp140.mut7.modSF162	(1951)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACTGGTTCGACATCAGCAA	
gp140.mut8.modSF162	(1951)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACTGGTTCGACATCAGCAA	
gp120.modSF162	(1513)	-----	
Consensus	(1951)	CTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACTGGTTCGACATCAGCAA	2001
			2050
gp160.modSF162	(2001)	GTGGCTGTGGTACATCAAGATCTTCATGATGCTGGGGCTGGTGG	
gp160.modSF162.de1V2	(1920)	GTGGCTGTGGTACATCAAGATCTTCATGATGCTGGGGCTGGTGG	
gp160.modSF162.de1V2	(1812)	GTGGCTGTGGTACATCAAGATCTTCATGATGCTGGGGCTGGTGG	
gp140.modSF162	(2001)	GTGGCTGTGGTACATCAACTCGAG-----	
gp140.mut.modSF162	(2001)	GTGGCTGTGGTACATCAACTCGAG-----	

FIG. 66A-14

93 / 131

gp140 .mut7 .modSF162	(2001)	GTCGGCTGTGGTACATCTAACTCGAG		2100
gp140 .mut8 .modSF162	(2001)	GTCGGCTGTGGTACATCTAACTCGAG		
gp120 .modSF162	(1513)			
Consensus	(2001)	GTGGCTGTGGTACATCTAACTCGAG		
	2051			
gp160 .modSF162	(2051)	GCCTGCGCATCGTGTACCCGTGCTAACCGTGTGAGCATCGTGAACCGCGTGGGCCAG		
gp160 .modSF162 .delV2	(1970)	GCCTGCGCATCGTGTACCCGTGCTAACCGTGTGAGCATCGTGAACCGCGTGGGCCAG		
gp160 .modSF162 .delV1V2	(1862)	GCCTGCGCATCGTGTACCCGTGCTAACCGTGTGAGCATCGTGAACCGCGTGGGCCAG		
gp140 .modSF162	(2026)			
gp140 .mut .modSF162	(2026)			
gp140 .mut7 .modSF162	(2026)			
gp140 .mut8 .modSF162	(2026)			
gp120 .modSF162	(1513)			
Consensus	(2051)	GTGGCTGTGGTACATCTAACTCGAG		
	2101			
gp160 .modSF162	(2101)	GGCTACAGCCCCCTGAGCTTCCAGACCCCGCTTCCAGACCCCGCTTCCCGCCCCCGGGCCC		
gp160 .modSF162 .delV2	(2020)	GGCTACAGCCCCCTGAGCTTCCAGACCCCGCTTCCAGACCCCGCTTCCCGCCCCCGGGCCC		
gp160 .modSF162 .delV1V2	(1912)	GGCTACAGCCCCCTGAGCTTCCAGACCCCGCTTCCCGCCCCCGGGCCC		
gp140 .modSF162	(2026)			
gp140 .mut .modSF162	(2026)			
gp140 .mut7 .modSF162	(2026)			
gp140 .mut8 .modSF162	(2026)			
gp120 .modSF162	(1513)			
Consensus	(2101)	GTGGCTGTGGTACATCTAACTCGAG		
	2151			
gp160 .modSF162	(2151)	CGACCCCCCGAGGGCATCGAGGAGGAGGGGGGAGCCGGGACGGGGGACCGGGGACCC		
gp160 .modSF162 .delV2	(2070)	CGACCCCCCGAGGGCATCGAGGAGGAGGGGGGAGCCGGGACGGGGGACCGGGGACCC		
gp160 .modSF162 .delV1V2	(1962)	CGACCCCCCGAGGGCATCGAGGAGGAGGGGGGAGCCGGGACGGGGGACCGGGGACCC		
gp140 .modSF162	(2026)			
gp140 .mut .modSF162	(2026)			
gp140 .mut7 .modSF162	(2026)			
gp140 .mut8 .modSF162	(2026)			
gp120 .modSF162	(1513)			
Consensus	(2101)	GTGGCTGTGGTACATCTAACTCGAG		
	2200			
gp160 .modSF162	(2200)	GTGGCTGTGGTACATCTAACTCGAG		
gp160 .modSF162 .delV2	(2120)	GTGGCTGTGGTACATCTAACTCGAG		
gp160 .modSF162 .delV1V2	(2012)	GTGGCTGTGGTACATCTAACTCGAG		
gp140 .modSF162	(1926)			
gp140 .mut .modSF162	(1926)			
gp140 .mut7 .modSF162	(1926)			
gp140 .mut8 .modSF162	(1926)			
gp120 .modSF162	(1513)			
Consensus	(2151)	GTGGCTGTGGTACATCTAACTCGAG		

FIG. 66A-15

94 / 131

		2201		2250
gp160.modSF162	(2201)	GCAGCAGCCCCCTGGTGCACGGCCTGGCTGGCCCTGATCTGGGACGGACTG		
gp160.modSF162.delV2	(2120)	CGAGCAGCCCCCTGGTGCACGGCCTGGCTGGCCCTGATCTGGGACGGACTG		
gp160.modSF162.delV1V2	(2012)	CGAGCAGCCCCCTGGTGCACGGCCTGGCTGGCCCTGATCTGGGACGGACTG		
gp140.modSF162	(2026)	-		
gp140.mut.modSF162	(2026)	-		
gp140.mut7.modSF162	(2026)	-		
gp140.mut8.modSF162	(2026)	-		
gp120.modSF162	(1513)	-		
Consensus	(2201)	2251		2300
gp160.modSF162	(2251)	CGCAGCCTGGCCCTGAGCTACCAACCGCTTGCGCGACCTGATCCTGAT		
gp160.modSF162.delV2	(2170)	CGCAGCCTGGCCCTGAGCTACCAACCGCTTGCGCGACCTGATCCTGAT		
gp160.modSF162.delV1V2	(2062)	CGCAGCCTGGCCCTGAGCTACCAACCGCTTGCGCGACCTGATCCTGAT		
gp140.modSF162	(2026)	-		
gp140.mut.modSF162	(2026)	-		
gp140.mut7.modSF162	(2026)	-		
gp140.mut8.modSF162	(2026)	-		
gp120.modSF162	(1513)	-		
Consensus	(2251)	2301		2350
gp160.modSF162	(2301)	CGCCGCCGGCATCGTGGAGCTGCTGGGCCGCCGGCTGGGAGGCCCTGA		
gp160.modSF162.delV2	(2220)	CGCCGCCGGCATCGTGGAGCTGCTGGGCCGCCGGCTGGGAGGCCCTGA		
gp160.modSF162.delV1V2	(2112)	CGCCGCCGGCATCGTGGAGCTGCTGGGCCGCCGGCTGGGAGGCCCTGA		

FIG. 66A-16

gp140.modSF162	(2026)	-	2351	2400
gp140.mut.modSF162	(2026)	-	(2351)	AGTACTGGGGCAACCTGCTGCAGTACTGGATCCAGGAGCTGAAGAACAGC
gp140.mut7.modSF162	(2026)	-	(2270)	AGTACTGGGGCAACCTGCTGCAGTACTGGATCCAGGAGCTGAAGAACAGC
gp140.mut8.modSF162	(2026)	-	(2162)	AGTACTGGGGCAACCTGCTGCAGTACTGGATCCAGGAGCTGAAGAACAGC
gp120.modSF162	(1513)	-	(2026)	-
Consensus	(2301)		(2351)	AGTACTGGGGCAACCTGCTGCAGTACTGGATCCAGGAGCTGAAGAACAGC
gp160.modSF162	(2026)	-	(2270)	AGTACTGGGGCAACCTGCTGCAGTACTGGATCCAGGAGCTGAAGAACAGC
gp160.modSF162.delV2	(2026)	-	(2162)	AGTACTGGGGCAACCTGCTGCAGTACTGGATCCAGGAGCTGAAGAACAGC
gp160.modSF162.delV1V2	(2026)	-	(2026)	-
gp140.modSF162	(2026)	-	(2212)	GCCGTGAGCCTGTTCGACGCCATCGCCATGCCCATGCCGAGGGCACCGA
gp140.mut.modSF162	(2026)	-	(2026)	-
gp140.mut7.modSF162	(2026)	-	(2026)	-
gp140.mut8.modSF162	(2026)	-	(2026)	-
gp120.modSF162	(1513)	-	(2026)	-
Consensus	(2301)		(2351)	AGTACTGGGGCAACCTGCTGCAGTACTGGATCCAGGAGCTGAAGAACAGC
gp160.modSF162	(2026)	-	(2401)	GCCGTGAGCCTGTTCGACGCCATCGCCATGCCGAGGGCACCGA
gp160.modSF162.delV2	(2026)	-	(2320)	GCCGTGAGCCTGTTCGACGCCATCGCCATGCCGAGGGCACCGA
gp160.modSF162.delV1V2	(2026)	-	(2212)	GCCGTGAGCCTGTTCGACGCCATCGCCATGCCGAGGGCACCGA
gp140.modSF162	(2026)	-	(2026)	-
gp140.mut.modSF162	(2026)	-	(2026)	-
gp140.mut7.modSF162	(2026)	-	(2026)	-
gp140.mut8.modSF162	(2026)	-	(2026)	-
gp120.modSF162	(1513)	-	(2026)	-
Consensus	(2401)		(2401)	GCCGTGAGCCTGTTCGACGCCATCGCCATGCCGAGGGCACCGA

FIG. 66A-17

96 / 131

2451	gp160.modSF162 gp160.modSF162.delV2 gp160.modSF162.delV1V2 gp140.modSF162 gp140.mut.modSF162 gp140.mut7.modSF162 gp140.mut8.modSF162 gp120.modSF162 Consensus	(2451) (2370) (2262) (2026) (2026) (2026) (2026) (1513) (2451)	2500 CGGCATCATCGAGGTGGCCCAAGGCCATGGCCGGCCTTCCTGCACATCC CGGCATCATCGAGGTGGCCCAAGGCCATGGCCGGCCTTCCTGCACATCC CGGCATCATCGAGGTGGCCCAAGGCCATGGCCGGCCTTCCTGCACATCC -
2501	gp160.modSF162 gp160.modSF162.delV2 gp160.modSF162.delV1V2 gp140.modSF162 gp140.mut.modSF162 gp140.mut7.modSF162 gp140.mut8.modSF162 gp120.modSF162 Consensus	(2501) (2420) (2312) (2026) (2026) (2026) (2026) (1513) (2501)	2547 CCGGCCGCATCCGCAGGGCTTCGAGGGCCCTGCTGTAACTCGAG CCGGCCGCATCCGCAGGGCTTCGAGGGCCCTGCTGTAACTCGAG CCGGCCGCATCCGCAGGGCTTCGAGGGCCCTGCTGTAACTCGAG -

FIG. 66A-18

	Start of tPA
gp160	1
gp160 del V1	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT
gp160 del V2	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT
gp160 del V1-2	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT
gp 160 del 128-194	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT
gp140TM	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT
gp140	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT
gp140mut	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT
gp120	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT
Consensus	(1) GAATTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCT
	40
gp160	41
gp160 del V1	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCAG
gp160 del V2	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCAG
gp160 del V1-2	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCAG
gp 160 del 128-194	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCAG
gp140TM	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCAG
gp140	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCAG
gp140mut	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCAG
gp120	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCAG
Consensus	(41) GTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCAG
	80
end of tPA	81
gp160	120
gp160 del V1	(81) CGCCACCACCGTGTGGTGACCGTGTACTACGGCGTG
gp160 del V2	(81) CGCCACCACCGTGTGGTGACCGTGTACTACGGCGTG
gp160 del V1-2	(81) CGCCACCACCGTGTGGTGACCGTGTACTACGGCGTG
gp 160 del 128-194	(81) CGCCACCACCGTGTGGTGACCGTGTACTACGGCGTG
gp140TM	(81) CGCCACCACCGTGTGGTGACCGTGTACTACGGCGTG
gp140	(81) CGCCACCACCGTGTGGTGACCGTGTACTACGGCGTG
gp140mut	(81) CGCCACCACCGTGTGGTGACCGTGTACTACGGCGTG
gp120	(81) CGCCACCACCGTGTGGTGACCGTGTACTACGGCGTG
Consensus	(81) CGCCACCACCGTGTGGTGACCGTGTACTACGGCGTG
	121
gp 160	160
gp160 del V1	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA
gp160 del V2	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA
gp160 del V1-2	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA
gp 160 del 128-194	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA
gp140TM	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA
gp140	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA
gp140mut	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA
gp120	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA
Consensus	(121) CCCGTGTGGAAGGAGGCCACCAACCCCTGTTCTGCGCCA

FIG. 66B-1

98 / 131

		161	200
gp160	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp160 del V1	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp160 del V2	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp160 del V1-2	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp 160 del 128-194	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp140TM	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp140	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp140mut	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
gp120	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
Consensus	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTG	
		201	240
gp160	(201)	GGCCACCCACGCCCTGCGTGCCCACCGACCCCAACCCCCAG	
gp160 del V1	(201)	GGCCACCCACGCCCTGCGTGCCCACCGACCCCAACCCCCAG	
gp160 del V2	(201)	GGCCACCCACGCCCTGCGTGCCCACCGACCCCAACCCCCAG	
gp160 del V1-2	(201)	GGCCACCCACGCCCTGCGTGCCCACCGACCCCAACCCCCAG	
gp 160 del 128-194	(201)	GGCCACCCACGCCCTGCGTGCCCACCGACCCCAACCCCCAG	
gp140TM	(201)	GGCCACCCACGCCCTGCGTGCCCACCGACCCCAACCCCCAG	
gp140	(201)	GGCCACCCACGCCCTGCGTGCCCACCGACCCCAACCCCCAG	
gp140mut	(201)	GGCCACCCACGCCCTGCGTGCCCACCGACCCCAACCCCCAG	
gp120	(201)	GGCCACCCACGCCCTGCGTGCCCACCGACCCCAACCCCCAG	
Consensus	(201)	GGCCACCCACGCCCTGCGTGCCCACCGACCCCAACCCCCAG	
		241	280
gp160	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT	
gp160 del V1	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT	
gp160 del V2	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT	
gp160 del V1-2	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT	
gp 160 del 128-194	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT	
gp140TM	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT	
gp140	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT	
gp140mut	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT	
gp120	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT	
Consensus	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCACATGT	
		281	320
gp160	(281)	GGAAGAACAAACATGGTGGACGAGATGCATGAGGACATCAT	
gp160 del V1	(281)	GGAAGAACAAACATGGTGGACGAGATGCATGAGGACATCAT	
gp160 del V2	(281)	GGAAGAACAAACATGGTGGACGAGATGCATGAGGACATCAT	
gp160 del V1-2	(281)	GGAAGAACAAACATGGTGGACGAGATGCATGAGGACATCAT	
gp 160 del 128-194	(281)	GGAAGAACAAACATGGTGGACGAGATGCATGAGGACATCAT	
gp140TM	(281)	GGAAGAACAAACATGGTGGACGAGATGCATGAGGACATCAT	
gp140	(281)	GGAAGAACAAACATGGTGGACGAGATGCATGAGGACATCAT	
gp140mut	(281)	GGAAGAACAAACATGGTGGACGAGATGCATGAGGACATCAT	
gp120	(281)	GGAAGAACAAACATGGTGGACGAGATGCATGAGGACATCAT	
Consensus	(281)	GGAAGAACAAACATGGTGGACGAGATGCATGAGGACATCAT	
		321	360
gp160	(321)	CAGCCTGTGGGACCAAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp160 del V1	(321)	CAGCCTGTGGGACCAAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp160 del V2	(321)	CAGCCTGTGGGACCAAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp160 del V1-2	(321)	CAGCCTGTGGGACCAAGAGCCTGAAGCCCTGCGTGGCGCC	
gp 160 del 128-194	(321)	CAGCCTGTGGGACCAAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp140TM	(321)	CAGCCTGTGGGACCAAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp140	(321)	CAGCCTGTGGGACCAAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp140mut	(321)	CAGCCTGTGGGACCAAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp120	(321)	CAGCCTGTGGGACCAAGAGCCTGAAGCCCTGCGTGAAGCTG	
Consensus	(321)	CAGCCTGTGGGACCAAGAGCCTGAAGCCCTGCGTGAAGCTG	

FIG. 66B-2

99 / 131

	361	400
gp160	(361) ACCCCCTGTGCGTACCCCTGAAC TGACCCGACAAGCTGA	
gp160 del V1	(361) ACCCCCTGTGCGTACCCCTGAAC TGACCCGACAAGCTGG	
gp160 del V2	(361) ACCCCCTGTGCGTACCCCTGAAC TGACCCGACAAGCTGA	
gp160 del V1-2	(361) GGC-----	
gp 160 del 128-194	(361) ACCCCCTGTGCGTGGGGCAGGG-----	
gp140TM	(361) ACCCCCTGTGCGTACCCCTGAAC TGACCCGACAAGCTGA	
gp140	(361) ACCCCCTGTGCGTACCCCTGAAC TGACCCGACAAGCTGA	
gp140mut	(361) ACCCCCTGTGCGTACCCCTGAAC TGACCCGACAAGCTGA	
gp120	(361) ACCCCCTGTGCGTACCCCTGAAC TGACCCGACAAGCTGA	
Consensus	(361) ACCCCCTGTGCGTACCCCTGAAC TGACCCGACAAGCTGA	
	401	440
gp160	(401) CCGGCAGCACCAACGGCACCAACAGCACCGACAGCGGCACCAA	
gp160 del V1	(401) GCGCCGGC-----	
gp160 del V2	(401) CCGGCAGCACCAACGGCACCAACAGCACCGACAGCGGCACCAA	
gp160 del V1-2	(364) -----	
gp 160 del 128-194	(385) -----	
gp140TM	(401) CCGGCAGCACCAACGGCACCAACAGCACCGACAGCGGCACCAA	
gp140	(401) CCGGCAGCACCAACGGCACCAACAGCACCGACAGCGGCACCAA	
gp140mut	(401) CCGGCAGCACCAACGGCACCAACAGCACCGACAGCGGCACCAA	
gp120	(401) CCGGCAGCACCAACGGCACCAACAGCACCGACAGCGGCACCAA	
Consensus	(401) CCGGCAGCACCAACGGCACCAACAGCACCGACAGCGGCACCAA	
	441	480
gp160	(441) CAGCACCAGCGGCACCAACAGCACCAACAGCACCC	
gp160 del V1	(409) -----	
gp160 del V2	(441) CAGCACCAGCGGCACCAACAGCACCAACAGCACCC	
gp160 del V1-2	(364) -----	
gp 160 del 128-194	(385) -----	
gp140TM	(441) CAGCACCAGCGGCACCAACAGCACCAACAGCACCC	
gp140	(441) CAGCACCAGCGGCACCAACAGCACCAACAGCACCC	
gp140mut	(441) CAGCACCAGCGGCACCAACAGCACCAACAGCACCC	
gp120	(441) CAGCACCAGCGGCACCAACAGCACCAACAGCACCC	
Consensus	(441) CAGCACCAGCGGCACCAACAGCACCAACAGCACCC	
	481	520
gp160	(481) GACAGCTGGGAGAAGATGCCGAGGGCAGATCAAGAACT	
gp160 del V1	(409) -----GGCGAGATCAAGAACT	
gp160 del V2	(481) GACAGCTGGGAGAAGATGCCGAGGGCAGATCAAGAACT	
gp160 del V1-2	(364) -----	
gp 160 del 128-194	(385) -----	
gp140TM	(481) GACAGCTGGGAGAAGATGCCGAGGGCAGATCAAGAACT	
gp140	(481) GACAGCTGGGAGAAGATGCCGAGGGCAGATCAAGAACT	
gp140mut	(481) GACAGCTGGGAGAAGATGCCGAGGGCAGATCAAGAACT	
gp120	(481) GACAGCTGGGAGAAGATGCCGAGGGCAGATCAAGAACT	
Consensus	(481) GACAGCTGGGAGAAGATGCCGAGGGCAGATCAAGAACT	
	521	560
gp160	(521) GCAGCTTCAACATCACCACCGCGTGGCGACAAGGTGCA	
gp160 del V1	(521) GCAGCTTCAACATCACCACCGCGTGGCGACAAGGTGCA	
gp160 del V2	(521) GCAGCTTCAACATCGGCGCCGC-----	
gp160 del V1-2	(521) -----	
gp 160 del 128-194	(521) -----	
gp140TM	(521) GCAGCTTCAACATCACCACCGCGTGGCGACAAGGTGCA	
gp140	(521) GCAGCTTCAACATCACCACCGCGTGGCGACAAGGTGCA	
gp140mut	(521) GCAGCTTCAACATCACCACCGCGTGGCGACAAGGTGCA	
gp120	(521) GCAGCTTCAACATCACCACCGCGTGGCGACAAGGTGCA	
Consensus	(521) GCAGCTTCAACATCACCACCGCGTGGCGACAAGGTGCA	

FIG. 66B-3

100 / 131

	561	600
gp160	(561) GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGC CCC	
gp160 del V1	(465) GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGC CCC	
gp160 del V2	(544) -----	
gp160 del V1-2	(364) -----	
gp 160 del 128-194	(385) -----	
gp140TM	(561) GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGC CCC	
gp140	(561) GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGC CCC	
gp140mut	(561) GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGC CCC	
gp120	(561) GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGC CCC	
Consensus	(561) GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGC CCC	
	601	640
gp160	(601) ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA	
gp160 del V1	(505) ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA	
gp160 del V2	(544) -----CGCCTGATCAACTGCA	
gp160 del V1-2	(364) -----	
gp 160 del 128-194	(385) -----AACTGCG	
gp140TM	(601) ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA	
gp140	(601) ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA	
gp140mut	(601) ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA	
gp120	(601) ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA	
Consensus	(601) ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA	
	641	680
gp160	(641) ACACCAGCGTGATCACCCAGGGCTGCCCAAGGTGAGCTT	
gp160 del V1	(545) ACACCAGCGTGATCACCCAGGGCTGCCCAAGGTGAGCTT	
gp160 del V2	(560) ACACCAGCGTGATCACCCAGGGCTGCCCAAGGTGAGCTT	
gp160 del V1-2	(364) -----CAGGGCTGCCCAAGGTGAGCTT	
gp 160 del 128-194	(392) AGACCAGCGTGATCACCCAGGGCTGCCCAAGGTGAGCTT	
gp140TM	(641) ACACCAGCGTGATCACCCAGGGCTGCCCAAGGTGAGCTT	
gp140	(641) ACACCAGCGTGATCACCCAGGGCTGCCCAAGGTGAGCTT	
gp140mut	(641) ACACCAGCGTGATCACCCAGGGCTGCCCAAGGTGAGCTT	
gp120	(641) ACACCAGCGTGATCACCCAGGGCTGCCCAAGGTGAGCTT	
Consensus	(641) ACACCAGCGTGATCACCCAGGGCTGCCCAAGGTGAGCTT	
	681	720
gp160	(681) CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC	
gp160 del V1	(585) CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC	
gp160 del V2	(600) CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC	
gp160 del V1-2	(387) CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC	
gp 160 del 128-194	(432) CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC	
gp140TM	(681) CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC	
gp140	(681) CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC	
gp140mut	(681) CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC	
gp120	(681) CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC	
Consensus	(681) CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC	
	721	760
gp160	(721) GCCATCCTGAAGTGAAGGACAAGAAGTTAACGGCACCG	
gp160 del V1	(625) GCCATCCTGAAGTGAAGGACAAGAAGAAAGTTAACGGCACCG	
gp160 del V2	(640) GCCATCCTGAAGTGAAGGACAAGAAGAAAGTTAACGGCACCG	
gp160 del V1-2	(427) GCCATCCTGAAGTGAAGGACAAGAAGAAAGTTAACGGCACCG	
gp 160 del 128-194	(472) GCCATCCTGAAGTGAAGGACAAGAAGAAAGTTAACGGCACCG	
gp140TM	(721) GCCATCCTGAAGTGAAGGACAAGAAGAAAGTTAACGGCACCG	
gp140	(721) GCCATCCTGAAGTGAAGGACAAGAAGAAAGTTAACGGCACCG	
gp140mut	(721) GCCATCCTGAAGTGAAGGACAAGAAGAAAGTTAACGGCACCG	
gp120	(721) GCCATCCTGAAGTGAAGGACAAGAAGAAAGTTAACGGCACCG	
Consensus	(721) GCCATCCTGAAGTGAAGGACAAGAAGAAAGTTAACGGCACCG	

FIG. 66B-4

101 / 131

		761	800
gp160	(761) GCCCCCTGCAAGAACGTGAGCACCCTGCAGTCACCCACGG		
gp160 del V1	(665) GCCCCCTGCAAGAACGTGAGCACCCTGCAGTCACCCACGG		
gp160 del V2	(680) GCCCCCTGCAAGAACGTGAGCACCCTGCAGTCACCCACGG		
gp160 del V1-2	(467) GCCCCCTGCAAGAACGTGAGCACCCTGCAGTCACCCACGG		
gp 160 del 128-194	(512) GCCCCCTGCAAGAACGTGAGCACCCTGCAGTCACCCACGG		
gp140TM	(761) GCCCCCTGCAAGAACGTGAGCACCCTGCAGTCACCCACGG		
gp140	(761) GCCCCCTGCAAGAACGTGAGCACCCTGCAGTCACCCACGG		
gp140mut	(761) GCCCCCTGCAAGAACGTGAGCACCCTGCAGTCACCCACGG		
gp120	(761) GCCCCCTGCAAGAACGTGAGCACCCTGCAGTCACCCACGG		
Consensus	(761) GCCCCCTGCAAGAACGTGAGCACCCTGCAGTCACCCACGG		
		801	840
gp160	(801) CATCCGGCCCCGTGGTGAGCACCAGCTGCTGCTGAACGGC		
gp160 del V1	(705) CATCCGGCCCCGTGGTGAGCACCAGCTGCTGCTGAACGGC		
gp160 del V2	(720) CATCCGGCCCCGTGGTGAGCACCAGCTGCTGCTGAACGGC		
gp160 del V1-2	(507) CATCCGGCCCCGTGGTGAGCACCAGCTGCTGCTGAACGGC		
gp 160 del 128-194	(552) CATCCGGCCCCGTGGTGAGCACCAGCTGCTGCTGAACGGC		
gp140TM	(801) CATCCGGCCCCGTGGTGAGCACCAGCTGCTGCTGAACGGC		
gp140	(801) CATCCGGCCCCGTGGTGAGCACCAGCTGCTGCTGAACGGC		
gp140mut	(801) CATCCGGCCCCGTGGTGAGCACCAGCTGCTGCTGAACGGC		
gp120	(801) CATCCGGCCCCGTGGTGAGCACCAGCTGCTGCTGAACGGC		
Consensus	(801) CATCCGGCCCCGTGGTGAGCACCAGCTGCTGCTGAACGGC		
		841	880
gp160	(841) AGCCTGGCCGAGGAGGAGATCGTGCCTCGAGAACT		
gp160 del V1	(745) AGCCTGGCCGAGGAGGAGATCGTGCCTCGAGAACT		
gp160 del V2	(760) AGCCTGGCCGAGGAGGAGATCGTGCCTCGAGAACT		
gp160 del V1-2	(547) AGCCTGGCCGAGGAGGAGATCGTGCCTCGAGAACT		
gp 160 del 128-194	(592) AGCCTGGCCGAGGAGGAGATCGTGCCTCGAGAACT		
gp140TM	(841) AGCCTGGCCGAGGAGGAGATCGTGCCTCGAGAACT		
gp140	(841) AGCCTGGCCGAGGAGGAGATCGTGCCTCGAGAACT		
gp140mut	(841) AGCCTGGCCGAGGAGGAGATCGTGCCTCGAGAACT		
gp120	(841) AGCCTGGCCGAGGAGGAGATCGTGCCTCGAGAACT		
Consensus	(841) AGCCTGGCCGAGGAGGAGATCGTGCCTCGAGAACT		
		881	920
gp160	(881) TCACCGACAACGCCAAGACCATCATCGTCAGCTGAACGA		
gp160 del V1	(785) TCACCGACAACGCCAAGACCATCATCGTCAGCTGAACGA		
gp160 del V2	(800) TCACCGACAACGCCAAGACCATCATCGTCAGCTGAACGA		
gp160 del V1-2	(587) TCACCGACAACGCCAAGACCATCATCGTCAGCTGAACGA		
gp 160 del 128-194	(632) TCACCGACAACGCCAAGACCATCATCGTCAGCTGAACGA		
gp140TM	(881) TCACCGACAACGCCAAGACCATCATCGTCAGCTGAACGA		
gp140	(881) TCACCGACAACGCCAAGACCATCATCGTCAGCTGAACGA		
gp140mut	(881) TCACCGACAACGCCAAGACCATCATCGTCAGCTGAACGA		
gp120	(881) TCACCGACAACGCCAAGACCATCATCGTCAGCTGAACGA		
Consensus	(881) TCACCGACAACGCCAAGACCATCATCGTCAGCTGAACGA		
		921	960
gp160	(921) GTCCGTGGAGATCAAATGCATCCGCCCCAACAACACAG		
gp160 del V1	(825) GTCCGTGGAGATCAAATGCATCCGCCCCAACAACACAG		
gp160 del V2	(840) GTCCGTGGAGATCAAATGCATCCGCCCCAACAACACAG		
gp160 del V1-2	(627) GTCCGTGGAGATCAAATGCATCCGCCCCAACAACACAG		
gp 160 del 128-194	(672) GTCCGTGGAGATCAAATGCATCCGCCCCAACAACACAG		
gp140TM	(921) GTCCGTGGAGATCAAATGCATCCGCCCCAACAACACAG		
gp140	(921) GTCCGTGGAGATCAAATGCATCCGCCCCAACAACACAG		
gp140mut	(921) GTCCGTGGAGATCAAATGCATCCGCCCCAACAACACAG		
gp120	(921) GTCCGTGGAGATCAAATGCATCCGCCCCAACAACACAG		
Consensus	(921) GTCCGTGGAGATCAAATGCATCCGCCCCAACAACACAG		

FIG. 66B-5

102 / 131

		961	1000
gp160	(961)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG	
gp160 del V1	(865)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG	
gp160 del V2	(880)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG	
gp160 del V1-2	(667)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG	
gp 160 del 128-194	(712)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG	
gp140TM	(961)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG	
gp140	(961)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG	
gp140mut	(961)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG	
gp120	(961)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG	
Consensus	(961)	CGTAAGAGCATCCACATCGGCCCCGGCCGCCCTTCTACG	
	1001		1040
gp160	(1001)	CCACCGGGCAGACATCATCGGCAGACATCCGCCAGGCCCCTACTG	
gp160 del V1	(905)	CCACCGGGCAGACATCATCGGCAGACATCCGCCAGGCCCCTACTG	
gp160 del V2	(920)	CCACCGGGCAGACATCATCGGCAGACATCCGCCAGGCCCCTACTG	
gp160 del V1-2	(707)	CCACCGGGCAGACATCATCGGCAGACATCCGCCAGGCCCCTACTG	
gp 160 del 128-194	(752)	CCACCGGGCAGACATCATCGGCAGACATCCGCCAGGCCCCTACTG	
gp140TM	(1001)	CCACCGGGCAGACATCATCGGCAGACATCCGCCAGGCCCCTACTG	
gp140	(1001)	CCACCGGGCAGACATCATCGGCAGACATCCGCCAGGCCCCTACTG	
gp140mut	(1001)	CCACCGGGCAGACATCATCGGCAGACATCCGCCAGGCCCCTACTG	
gp120	(1001)	CCACCGGGCAGACATCATCGGCAGACATCCGCCAGGCCCCTACTG	
Consensus	(1001)	CCACCGGGCAGACATCATCGGCAGACATCCGCCAGGCCCCTACTG	
	1041		1080
gp160	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp160 del V1	(945)	CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp160 del V2	(960)	CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp160 del V1-2	(747)	CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp 160 del 128-194	(792)	CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp140TM	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp140	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp140mut	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp120	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
Consensus	(1041)	CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
	1081		1120
gp160	(1081)	ATCGTGGAGAACGCTCGCGAGCAGTCGGCAACAACAAGA	
gp160 del V1	(985)	ATCGTGGAGAACGCTCGCGAGCAGTCGGCAACAACAAGA	
gp160 del V2	(1000)	ATCGTGGAGAACGCTCGCGAGCAGTCGGCAACAACAAGA	
gp160 del V1-2	(787)	ATCGTGGAGAACGCTCGCGAGCAGTCGGCAACAACAAGA	
gp 160 del 128-194	(832)	ATCGTGGAGAACGCTCGCGAGCAGTCGGCAACAACAAGA	
gp140TM	(1081)	ATCGTGGAGAACGCTCGCGAGCAGTCGGCAACAACAAGA	
gp140	(1081)	ATCGTGGAGAACGCTCGCGAGCAGTCGGCAACAACAAGA	
gp140mut	(1081)	ATCGTGGAGAACGCTCGCGAGCAGTCGGCAACAACAAGA	
gp120	(1081)	ATCGTGGAGAACGCTCGCGAGCAGTCGGCAACAACAAGA	
Consensus	(1081)	ATCGTGGAGAACGCTCGCGAGCAGTCGGCAACAACAAGA	
	1121		1160
gp160	(1121)	CCATCATTTCAACAGCAGCAGCGGGGGGACCCCGAGAT	
gp160 del V1	(1025)	CCATCATTTCAACAGCAGCAGCGGGGGGACCCCGAGAT	
gp160 del V2	(1040)	CCATCATTTCAACAGCAGCAGCGGGGGGACCCCGAGAT	
gp160 del V1-2	(827)	CCATCATTTCAACAGCAGCAGCGGGGGGACCCCGAGAT	
gp 160 del 128-194	(872)	CCATCATTTCAACAGCAGCAGCGGGGGGACCCCGAGAT	
gp140TM	(1121)	CCATCATTTCAACAGCAGCAGCGGGGGGACCCCGAGAT	
gp140	(1121)	CCATCATTTCAACAGCAGCAGCGGGGGGACCCCGAGAT	
gp140mut	(1121)	CCATCATTTCAACAGCAGCAGCGGGGGGACCCCGAGAT	
gp120	(1121)	CCATCATTTCAACAGCAGCAGCGGGGGGACCCCGAGAT	
Consensus	(1121)	CCATCATTTCAACAGCAGCAGCGGGGGGACCCCGAGAT	

FIG. 66B-6

103 / 131

		1161	1200
gp160	(1161)	CGTGTCCACAGCTCAACTGCGCGGCAGTTCTTCTAC	
gp160 del V1	(1065)	CGTGTCCACAGCTCAACTGCGCGGCAGTTCTTCTAC	
gp160 del V2	(1080)	CGTGTCCACAGCTCAACTGCGCGGCAGTTCTTCTAC	
gp160 del V1-2	(867)	CGTGTCCACAGCTCAACTGCGCGGCAGTTCTTCTAC	
gp 160 del 128-194	(912)	CGTGTCCACAGCTCAACTGCGCGGCAGTTCTTCTAC	
gp140TM	(1161)	CGTGTCCACAGCTCAACTGCGCGGCAGTTCTTCTAC	
gp140	(1161)	CGTGTCCACAGCTCAACTGCGCGGCAGTTCTTCTAC	
gp140mut	(1161)	CGTGTCCACAGCTCAACTGCGCGGCAGTTCTTCTAC	
gp120	(1161)	CGTGTCCACAGCTCAACTGCGCGGCAGTTCTTCTAC	
Consensus	(1161)	CGTGTCCACAGCTCAACTGCGCGGCAGTTCTTCTAC	
	1201		1240
gp160	(1201)	TGCAACACCAGCCAGCTGTTAACAGCACCTGGAACATCA	
gp160 del V1	(1105)	TGCAACACCAGCCAGCTGTTAACAGCACCTGGAACATCA	
gp160 del V2	(1120)	TGCAACACCAGCCAGCTGTTAACAGCACCTGGAACATCA	
gp160 del V1-2	(907)	TGCAACACCAGCCAGCTGTTAACAGCACCTGGAACATCA	
gp 160 del 128-194	(952)	TGCAACACCAGCCAGCTGTTAACAGCACCTGGAACATCA	
gp140TM	(1201)	TGCAACACCAGCCAGCTGTTAACAGCACCTGGAACATCA	
gp140	(1201)	TGCAACACCAGCCAGCTGTTAACAGCACCTGGAACATCA	
gp140mut	(1201)	TGCAACACCAGCCAGCTGTTAACAGCACCTGGAACATCA	
gp120	(1201)	TGCAACACCAGCCAGCTGTTAACAGCACCTGGAACATCA	
Consensus	(1201)	TGCAACACCAGCCAGCTGTTAACAGCACCTGGAACATCA	
	1241		1280
gp160	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT	
gp160 del V1	(1145)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT	
gp160 del V2	(1160)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT	
gp160 del V1-2	(947)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT	
gp 160 del 128-194	(992)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT	
gp140TM	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT	
gp140	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT	
gp140mut	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT	
gp120	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT	
Consensus	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT	
	1281		1320
gp160	(1281)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp160 del V1	(1185)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp160 del V2	(1200)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp160 del V1-2	(987)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp 160 del 128-194	(1032)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp140TM	(1281)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp140	(1281)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp140mut	(1281)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
gp120	(1281)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
Consensus	(1281)	CCTGCCCTGCCGATCCGCCAGATCATCAACATGTGGCAG	
	1321		1360
gp160	(1321)	GAGGTGGCAAGGCCATGTACGCCCCCCCCATCCGGGCC	
gp160 del V1	(1225)	GAGGTGGCAAGGCCATGTACGCCCCCCCCATCCGGGCC	
gp160 del V2	(1240)	GAGGTGGCAAGGCCATGTACGCCCCCCCCATCCGGGCC	
gp160 del V1-2	(1027)	GAGGTGGCAAGGCCATGTACGCCCCCCCCATCCGGGCC	
gp 160 del 128-194	(1072)	GAGGTGGCAAGGCCATGTACGCCCCCCCCATCCGGGCC	
gp140TM	(1321)	GAGGTGGCAAGGCCATGTACGCCCCCCCCATCCGGGCC	
gp140	(1321)	GAGGTGGCAAGGCCATGTACGCCCCCCCCATCCGGGCC	
gp140mut	(1321)	GAGGTGGCAAGGCCATGTACGCCCCCCCCATCCGGGCC	
gp120	(1321)	GAGGTGGCAAGGCCATGTACGCCCCCCCCATCCGGGCC	
Consensus	(1321)	GAGGTGGCAAGGCCATGTACGCCCCCCCCATCCGGGCC	

FIG. 66B-7

104 / 131

			1400
gp160	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp160 del V1	(1265)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp160 del V2	(1280)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp160 del V1-2	(1067)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp 160 del 128-194	(1112)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp140TM	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp140	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp140mut	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp120	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
Consensus	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
		1440	
gp160	(1401)	CCCGCACGGCGGGACCAAACAACCGCACCAACGACACC	
gp160 del V1	(1305)	CCCGCACGGCGGGACCAAACAACCGCACCAACGACACC	
gp160 del V2	(1320)	CCCGCACGGCGGGACCAAACAACCGCACCAACGACACC	
gp160 del V1-2	(1107)	CCCGCACGGCGGGACCAAACAACCGCACCAACGACACC	
gp 160 del 128-194	(1152)	CCCGCACGGCGGGACCAAACAACCGCACCAACGACACC	
gp140TM	(1401)	CCCGCACGGCGGGACCAAACAACCGCACCAACGACACC	
gp140	(1401)	CCCGCACGGCGGGACCAAACAACCGCACCAACGACACC	
gp140mut	(1401)	CCCGCACGGCGGGACCAAACAACCGCACCAACGACACC	
gp120	(1401)	CCCGCACGGCGGGACCAAACAACCGCACCAACGACACC	
Consensus	(1401)	CCCGCACGGCGGGACCAAACAACCGCACCAACGACACC	
		1480	
gp160	(1441)	GAGACCTTCCGCCCCGGCGGGCAACATGAAGGACAAC	
gp160 del V1	(1345)	GAGACCTTCCGCCCCGGCGGGCAACATGAAGGACAAC	
gp160 del V2	(1360)	GAGACCTTCCGCCCCGGCGGGCAACATGAAGGACAAC	
gp160 del V1-2	(1147)	GAGACCTTCCGCCCCGGCGGGCAACATGAAGGACAAC	
gp 160 del 128-194	(1192)	GAGACCTTCCGCCCCGGCGGGCAACATGAAGGACAAC	
gp140TM	(1441)	GAGACCTTCCGCCCCGGCGGGCAACATGAAGGACAAC	
gp140	(1441)	GAGACCTTCCGCCCCGGCGGGCAACATGAAGGACAAC	
gp140mut	(1441)	GAGACCTTCCGCCCCGGCGGGCAACATGAAGGACAAC	
gp120	(1441)	GAGACCTTCCGCCCCGGCGGGCAACATGAAGGACAAC	
Consensus	(1441)	GAGACCTTCCGCCCCGGCGGGCAACATGAAGGACAAC	
		1520	
gp160	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp160 del V1	(1385)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp160 del V2	(1400)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp160 del V1-2	(1187)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp 160 del 128-194	(1232)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp140TM	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp140	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp140mut	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp120	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
Consensus	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
		1560	
gp160	(1521)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp160 del V1	(1425)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp160 del V2	(1440)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp160 del V1-2	(1227)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp 160 del 128-194	(1272)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp140TM	(1521)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp140	(1521)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp140mut	(1521)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
gp120	(1521)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	
Consensus	(1521)	GCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGGCCCGCGTG	

FIG. 66B-8

105 / 131

			1600
gp160	(1561)	GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
gp160 del V1	(1465)	GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
gp160 del V2	(1480)	GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
gp160 del V1-2	(1267)	GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
gp 160 del 128-194	(1312)	GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
gp140TM	(1561)	GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
gp140	(1561)	GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
gp140mut	(1561)	GTGCAGCGCGAGAAGAGCGCCGTGGGCCTGGGCGCCCTGT	
gp120	(1561)	GTGCAGCGCGAGAAGCGCTAAG-----	
Consensus	(1561)	GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
		1640	
gp160	(1601)	TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp160 del V1	(1505)	TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp160 del V2	(1520)	TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp160 del V1-2	(1307)	TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp 160 del 128-194	(1352)	TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp140TM	(1601)	TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp140	(1601)	TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp140mut	(1601)	TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp120	(1583)	ATATCGGATCCTCTAGA-----	
Consensus	(1601)	TCATCGGCTTCNTGGGCGCCGCCGGGAGCACCATGGGCG	
		1680	
gp160	(1640)	CCGCCTCCGTGACCCCTGACCGTGCAGGGCCCGCCAGCTGCT	
gp160 del V1	(1544)	CCGCCTCCGTGACCCCTGACCGTGCAGGGCCCGCCAGCTGCT	
gp160 del V2	(1559)	CCGCCTCCGTGACCCCTGACCGTGCAGGGCCCGCCAGCTGCT	
gp160 del V1-2	(1346)	CCGCCTCCGTGACCCCTGACCGTGCAGGGCCCGCCAGCTGCT	
gp 160 del 128-194	(1391)	CCGCCTCCGTGACCCCTGACCGTGCAGGGCCCGCCAGCTGCT	
gp140TM	(1640)	CCGCCTCCGTGACCCCTGACCGTGCAGGGCCCGCCAGCTGCT	
gp140	(1640)	CCGCCTCCGTGACCCCTGACCGTGCAGGGCCCGCCAGCTGCT	
gp140mut	(1640)	CCGCCTCCGTGACCCCTGACCGTGCAGGGCCCGCCAGCTGCT	
gp120	(1600)	-----	
Consensus	(1641)	CCGCCTCCGTGACCCCTGACCGTGCAGGGCCCGCCAGCTGCT	
		1720	
gp160	(1680)	GAGCGGCATCGTGCAGCAGCAGAACAACTGCTGCGCGCC	
gp160 del V1	(1584)	GAGCGGCATCGTGCAGCAGCAGAACAACTGCTGCGCGCC	
gp160 del V2	(1599)	GAGCGGCATCGTGCAGCAGCAGAACAACTGCTGCGCGCC	
gp160 del V1-2	(1386)	GAGCGGCATCGTGCAGCAGCAGAACAACTGCTGCGCGCC	
gp 160 del 128-194	(1431)	GAGCGGCATCGTGCAGCAGCAGAACAACTGCTGCGCGCC	
gp140TM	(1680)	GAGCGGCATCGTGCAGCAGCAGAACAACTGCTGCGCGCC	
gp140	(1680)	GAGCGGCATCGTGCAGCAGCAGAACAACTGCTGCGCGCC	
gp140mut	(1680)	GAGCGGCATCGTGCAGCAGCAGAACAACTGCTGCGCGCC	
gp120	(1600)	-----	
Consensus	(1681)	GAGCGGCATCGTGCAGCAGCAGAACAACTGCTGCGCGCC	
		1760	
gp160	(1720)	ATCGAGGCCACCGACCTGCTGCAGCTGACCGTGTGGG	
gp160 del V1	(1624)	ATCGAGGCCACCGACCTGCTGCAGCTGACCGTGTGGG	
gp160 del V2	(1639)	ATCGAGGCCACCGACCTGCTGCAGCTGACCGTGTGGG	
gp160 del V1-2	(1426)	ATCGAGGCCACCGACCTGCTGCAGCTGACCGTGTGGG	
gp 160 del 128-194	(1471)	ATCGAGGCCACCGACCTGCTGCAGCTGACCGTGTGGG	
gp140TM	(1720)	ATCGAGGCCACCGACCTGCTGCAGCTGACCGTGTGGG	
gp140	(1720)	ATCGAGGCCACCGACCTGCTGCAGCTGACCGTGTGGG	
gp140mut	(1720)	ATCGAGGCCACCGACCTGCTGCAGCTGACCGTGTGGG	
gp120	(1600)	-----	
Consensus	(1721)	ATCGAGGCCACCGACCTGCTGCAGCTGACCGTGTGGG	

FIG. 66B-9

106/131

		1761	1800
gp160	(1760)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG	
gp160 del V1	(1664)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG	
gp160 del V2	(1679)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG	
gp160 del V1-2	(1466)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG	
gp 160 del 128-194	(1511)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG	
gp140TM	(1760)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG	
gp140	(1760)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG	
gp140mut	(1760)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG	
gp120	(1600)	-----	
Consensus	(1761)	GCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG	
	1801	1840	
gp160	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp160 del V1	(1704)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp160 del V2	(1719)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp160 del V1-2	(1506)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp 160 del 128-194	(1551)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp140TM	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp140	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp140mut	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp120	(1600)	-----	
Consensus	(1801)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
	1841	1880	
gp160	(1840)	AGCGGCAAGCTGATCTGACCAACCACCGTGCCTGGAAACA	
gp160 del V1	(1744)	AGCGGCAAGCTGATCTGACCAACCACCGTGCCTGGAAACA	
gp160 del V2	(1759)	AGCGGCAAGCTGATCTGACCAACCACCGTGCCTGGAAACA	
gp160 del V1-2	(1546)	AGCGGCAAGCTGATCTGACCAACCACCGTGCCTGGAAACA	
gp 160 del 128-194	(1591)	AGCGGCAAGCTGATCTGACCAACCACCGTGCCTGGAAACA	
gp140TM	(1840)	AGCGGCAAGCTGATCTGACCAACCACCGTGCCTGGAAACA	
gp140	(1840)	AGCGGCAAGCTGATCTGACCAACCACCGTGCCTGGAAACA	
gp140mut	(1840)	AGCGGCAAGCTGATCTGACCAACCACCGTGCCTGGAAACA	
gp120	(1600)	-----	
Consensus	(1841)	AGCGGCAAGCTGATCTGACCAACCACCGTGCCTGGAAACA	
	1881	1920	
gp160	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp160 del V1	(1784)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp160 del V2	(1799)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp160 del V1-2	(1586)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp 160 del 128-194	(1631)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp140TM	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp140	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp140mut	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp120	(1600)	-----	
Consensus	(1881)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
	1921	1960	
gp160	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp160 del V1	(1824)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp160 del V2	(1839)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp160 del V1-2	(1626)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp 160 del 128-194	(1671)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp140TM	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp140	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp140mut	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp120	(1600)	-----	
Consensus	(1921)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	

FIG. 66B-10

107 / 131

		1961	2000
gp160	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp160 del V1	(1864)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp160 del V2	(1879)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp160 del V1-2	(1666)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp 160 del 128-194	(1711)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp140TM	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp140	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp140mut	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	
gp120	(1600)	-----	
Consensus	(1961)	ACCGGCCTGATCTACAACCTGATCGAGATGCCAGAAC	2040
	2001		
gp160	(2000)	AGCAGGAGAAGAACGGCAGGAGCTGCTGGAGCTGGACAA	
gp160 del V1	(1904)	AGCAGGAGAAGAACGGCAGGAGCTGCTGGAGCTGGACAA	
gp160 del V2	(1919)	AGCAGGAGAAGAACGGCAGGAGCTGCTGGAGCTGGACAA	
gp160 del V1-2	(1706)	AGCAGGAGAAGAACGGCAGGAGCTGCTGGAGCTGGACAA	
gp 160 del 128-194	(1751)	AGCAGGAGAAGAACGGCAGGAGCTGCTGGAGCTGGACAA	
gp140TM	(2000)	AGCAGGAGAAGAACGGCAGGAGCTGCTGGAGCTGGACAA	
gp140	(2000)	AGCAGGAGAAGAACGGCAGGAGCTGCTGGAGCTGGACAA	
gp140mut	(2000)	AGCAGGAGAAGAACGGCAGGAGCTGCTGGAGCTGGACAA	
gp120	(1600)	-----	
Consensus	(2001)	AGCAGGAGAAGAACGGCAGGAGCTGCTGGAGCTGGACAA	2080
	2041		
gp160	(2040)	GTGGGCCAGCCTGTGAACTGGTTCGACATCACCAACTGG	
gp160 del V1	(1944)	GTGGGCCAGCCTGTGAACTGGTTCGACATCACCAACTGG	
gp160 del V2	(1959)	GTGGGCCAGCCTGTGAACTGGTTCGACATCACCAACTGG	
gp160 del V1-2	(1746)	GTGGGCCAGCCTGTGAACTGGTTCGACATCACCAACTGG	
gp 160 del 128-194	(1791)	GTGGGCCAGCCTGTGAACTGGTTCGACATCACCAACTGG	
gp140TM	(2040)	GTGGGCCAGCCTGTGAACTGGTTCGACATCACCAACTGG	
gp140	(2040)	GTGGGCCAGCCTGTGAACTGGTTCGACATCACCAACTGG	
gp140mut	(2040)	GTGGGCCAGCCTGTGAACTGGTTCGACATCACCAACTGG	
gp120	(1600)	-----	
Consensus	(2041)	GTGGGCCAGCCTGTGAACTGGTTCGACATCACCAACTGG	2120
	2081		
gp160	(2080)	CTGTGGTACATCCGATTTCATCATGATCGTGGCGGCC	
gp160 del V1	(1984)	CTGTGGTACATCCGATTTCATCATGATCGTGGCGGCC	
gp160 del V2	(1999)	CTGTGGTACATCCGATTTCATCATGATCGTGGCGGCC	
gp160 del V1-2	(1786)	CTGTGGTACATCCGATTTCATCATGATCGTGGCGGCC	
gp 160 del 128-194	(1831)	CTGTGGTACATCCGATTTCATCATGATCGTGGCGGCC	
gp140TM	(2080)	CTGTGGTACATCCGATTTCATCATGATCGTGGCGGCC	
gp140	(2080)	CTGTGGTACATC-----	
gp140mut	(2080)	CTGTGGTACATC-----	
gp120	(1600)	-----	
Consensus	(2081)	CTGTGGTACATCCGATTTCATCATGATCGTGGCGGCC	2160
	2121		
gp160	(2120)	TGATCGGCCTGCGCATCGTGGTCCCGTGCTGAGCA-----	
gp160 del V1	(2024)	TGATCGGCCTGCGCATCGTGGTCCCGTGCTGAGCA-----	
gp160 del V2	(2039)	TGATCGGCCTGCGCATCGTGGTCCCGTGCTGAGCA-----	
gp160 del V1-2	(1826)	TGATCGGCCTGCGCATCGTGGTCCCGTGCTGAGCA-----	
gp 160 del 128-194	(1871)	TGATCGGCCTGCGCATCGTGGTCCCGTGCTGAGCA-----	
gp140TM	(2120)	TGATCGGCCTGCGCATCGTGGTCCCGTGCTGAGCATCGT	
gp140	(2092)	-----	
gp140mut	(2092)	-----	
gp120	(1600)	-----	
Consensus	(2121)	TGATCGGCCTGCGCATCGTGGTCCCGTGCTGAGCANNNN	

FIG. 66B-11

SUBSTITUTE SHEET (RULE 26)

108 / 131

			2161	2200
gp160	(2156)	-TCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC		
gp160 del V1	(2060)	-TCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC		
gp160 del V2	(2075)	-TCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC		
gp160 del V1-2	(1862)	-TCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC		
gp 160 del 128-194	(1907)	-TCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC		
gp140TM	(2160)	GTAAAGATATCGGATCCTCTAGA-----		
gp140	(2092)	-TAAGATATCGGATCCTCTAGA-----		
gp140mut	(2092)	-TAAGATATCGGATCCTCTAGA-----		
gp120	(1600)	-----		
Consensus	(2161)	NTCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCC	2240	
	2201			
gp160	(2195)	TGCAGACCCGCCTGCCGCCAGCGCGGGCCCCGACCGCCC		
gp160 del V1	(2099)	TGCAGACCCGCCTGCCGCCAGCGCGGGCCCCGACCGCCC		
gp160 del V2	(2114)	TGCAGACCCGCCTGCCGCCAGCGCGGGCCCCGACCGCCC		
gp160 del V1-2	(1901)	TGCAGACCCGCCTGCCGCCAGCGCGGGCCCCGACCGCCC		
gp 160 del 128-194	(1946)	TGCAGACCCGCCTGCCGCCAGCGCGGGCCCCGACCGCCC		
gp140TM	(2182)	-----		
gp140	(2113)	-----		
gp140mut	(2113)	-----		
gp120	(1600)	-----		
Consensus	(2201)	TGCAGACCCGCCTGCCGCCAGCGCGGGCCCCGACCGCCC	2280	
	2241			
gp160	(2235)	CGAGGGCATCGAGGAGGAGGGCGGCAGCGCGACCGCGAC		
gp160 del V1	(2139)	CGAGGGCATCGAGGAGGAGGGCGGCAGCGCGACCGCGAC		
gp160 del V2	(2154)	CGAGGGCATCGAGGAGGAGGGCGGCAGCGCGACCGCGAC		
gp160 del V1-2	(1941)	CGAGGGCATCGAGGAGGAGGGCGGCAGCGCGACCGCGAC		
gp 160 del 128-194	(1986)	CGAGGGCATCGAGGAGGAGGGCGGCAGCGCGACCGCGAC		
gp140TM	(2182)	-----		
gp140	(2113)	-----		
gp140mut	(2113)	-----		
gp120	(1600)	-----		
Consensus	(2241)	CGAGGGCATCGAGGAGGAGGGCGGCAGCGCGACCGCGAC	2320	
	2281			
gp160	(2275)	CGCAGCAACCGCCTGGTGCACGGCCTGCTGGCCCTGATCT		
gp160 del V1	(2179)	CGCAGCAACCGCCTGGTGCACGGCCTGCTGGCCCTGATCT		
gp160 del V2	(2194)	CGCAGCAACCGCCTGGTGCACGGCCTGCTGGCCCTGATCT		
gp160 del V1-2	(1981)	CGCAGCAACCGCCTGGTGCACGGCCTGCTGGCCCTGATCT		
gp 160 del 128-194	(2026)	CGCAGCAACCGCCTGGTGCACGGCCTGCTGGCCCTGATCT		
gp140TM	(2182)	-----		
gp140	(2113)	-----		
gp140mut	(2113)	-----		
gp120	(1600)	-----		
Consensus	(2281)	CGCAGCAACCGCCTGGTGCACGGCCTGCTGGCCCTGATCT	2360	
	2321			
gp160	(2315)	GGGACGACCTGCGCAGCCTGTGCCTGTTCAGCTACCACCG		
gp160 del V1	(2219)	GGGACGACCTGCGCAGCCTGTGCCTGTTCAGCTACCACCG		
gp160 del V2	(2234)	GGGACGACCTGCGCAGCCTGTGCCTGTTCAGCTACCACCG		
gp160 del V1-2	(2021)	GGGACGACCTGCGCAGCCTGTGCCTGTTCAGCTACCACCG		
gp 160 del 128-194	(2066)	GGGACGACCTGCGCAGCCTGTGCCTGTTCAGCTACCACCG		
gp140TM	(2182)	-----		
gp140	(2113)	-----		
gp140mut	(2113)	-----		
gp120	(1600)	-----		
Consensus	(2321)	GGGACGACCTGCGCAGCCTGTGCCTGTTCAGCTACCACCG		

FIG. 66B-12

SUBSTITUTE SHEET (RULE 26)

109/131

2361

2400

	gp160	(2355)	CCTGCGCACCTGCTGATCGTGGCCCGCATCGTGGAG
	gp160 del V1	(2259)	CCTGCGCACCTGCTGATCGTGGCCCGCATCGTGGAG
	gp160 del V2	(2274)	CCTGCGCACCTGCTGATCGTGGCCCGCATCGTGGAG
	gp160 del V1-2	(2061)	CCTGCGCACCTGCTGATCGTGGCCCGCATCGTGGAG
gp 160 del 128-194		(2106)	CCTGCGCACCTGCTGATCGTGGCCCGCATCGTGGAG
	gp140TM	(2182)	-----
	gp140	(2113)	-----
	gp140mut	(2113)	-----
	gp120	(1600)	-----
	Consensus	(2361)	CCTGCGCACCTGCTGATCGTGGCCCGCATCGTGGAG
			2440
		2401	
	gp160	(2395)	CTGCTGGGCCGCCGGCTGGGAGGCCCTGAAGTACTGGT
	gp160 del V1	(2299)	CTGCTGGGCCGCCGGCTGGGAGGCCCTGAAGTACTGGT
	gp160 del V2	(2314)	CTGCTGGGCCGCCGGCTGGGAGGCCCTGAAGTACTGGT
	gp160 del V1-2	(2101)	CTGCTGGGCCGCCGGCTGGGAGGCCCTGAAGTACTGGT
gp 160 del 128-194		(2146)	CTGCTGGGCCGCCGGCTGGGAGGCCCTGAAGTACTGGT
	gp140TM	(2182)	-----
	gp140	(2113)	-----
	gp140mut	(2113)	-----
	gp120	(1600)	-----
	Consensus	(2401)	CTGCTGGGCCGCCGGCTGGGAGGCCCTGAAGTACTGGT
			2480
		2441	
	gp160	(2435)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG
	gp160 del V1	(2339)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG
	gp160 del V2	(2354)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG
	gp160 del V1-2	(2141)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG
gp 160 del 128-194		(2186)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG
	gp140TM	(2182)	-----
	gp140	(2113)	-----
	gp140mut	(2113)	-----
	gp120	(1600)	-----
	Consensus	(2441)	GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG
			2520
		2481	
	gp160	(2475)	CGCCGTGAGCCTGTTCAACGCCACCGCCATGCCGTGGCC
	gp160 del V1	(2379)	CGCCGTGAGCCTGTTCAACGCCACCGCCATGCCGTGGCC
	gp160 del V2	(2394)	CGCCGTGAGCCTGTTCAACGCCACCGCCATGCCGTGGCC
	gp160 del V1-2	(2181)	CGCCGTGAGCCTGTTCAACGCCACCGCCATGCCGTGGCC
gp 160 del 128-194		(2226)	CGCCGTGAGCCTGTTCAACGCCACCGCCATGCCGTGGCC
	gp140TM	(2182)	-----
	gp140	(2113)	-----
	gp140mut	(2113)	-----
	gp120	(1600)	-----
	Consensus	(2481)	CGCCGTGAGCCTGTTCAACGCCACCGCCATGCCGTGGCC
			2560
		2521	
	gp160	(2515)	GAGGGCACCGACCGCATCATCGAGATCGTCAGCGCATCT
	gp160 del V1	(2419)	GAGGGCACCGACCGCATCATCGAGATCGTCAGCGCATCT
	gp160 del V2	(2434)	GAGGGCACCGACCGCATCATCGAGATCGTCAGCGCATCT
	gp160 del V1-2	(2221)	GAGGGCACCGACCGCATCATCGAGATCGTCAGCGCATCT
gp 160 del 128-194		(2266)	GAGGGCACCGACCGCATCATCGAGATCGTCAGCGCATCT
	gp140TM	(2182)	-----
	gp140	(2113)	-----
	gp140mut	(2113)	-----
	gp120	(1600)	-----
	Consensus	(2521)	GAGGGCACCGACCGCATCATCGAGATCGTCAGCGCATCT

FIG. 66B-13

SUBSTITUTE SHEET (RULE 26)

110 / 131

FIG. 66B-14

111 / 131

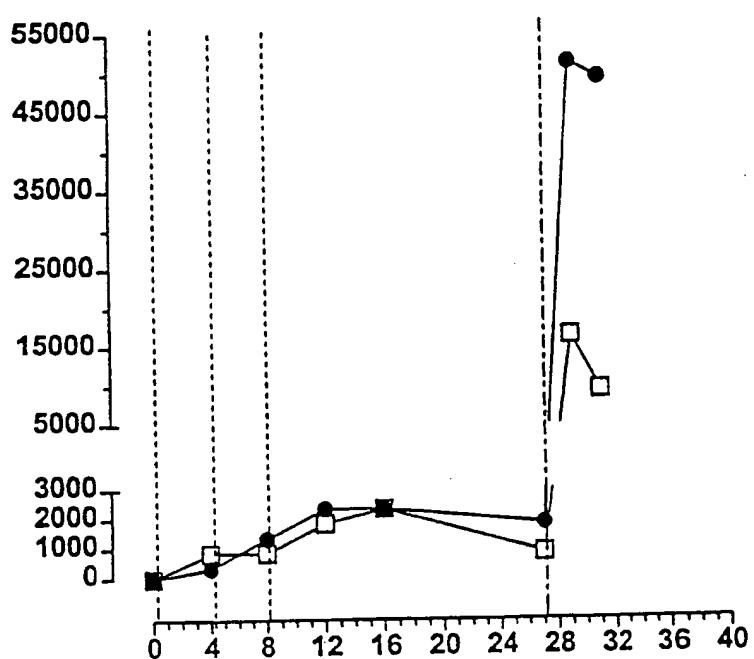


FIG. 67

112 / 131

HIV-1SF2 wt RT (PISPIET-->GIRKVL)

CCCATAGCCTATTGAAACTGTACAGTAAAATTAAAGCCAGGAATGGATGCCAAAAA
GTTAACGAAATGCCATTGACAGAAGAAAAATAAGCATTAGTAGAGATATGTACAGAA
ATGGAAAAGGAAGGGAAAATTCAAAAATTGGCCTGAAAATCCATAACAATACTCCAGTA
TTTGCTATAAAGAAAAAGACAGTACTAAATGGAGAAAATAGTAGATTTCAGAGAACTT
AATAAAAGAACTCAAGACTCTGGGAAGTTCACTTACCAACACCCCCGCAGGGTTA
AAAAAGAAAAATCAGTAACAGTATTGGATGTGGGTGATGCATACTTTCACTTCCTTA
GATAAAAGACTTTAGAAAGTATACTGCATTTACCATACCTAGTATAAACATGAGACACCA
GGGATTAGATATCAGTACAATGTGCTGCCACAGGGATGGAAAGGATCACCAGCAATATTC
CAAAGTAGCATGACAAAATCTTAGAGCCTTTAGAAAACAGAATCCAGACATAGTTATC
TATCAatacatggatTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAAGAAC
AAAATAGAGGAACGTGAGACAGCATTCTGAGGTTGGGATTACACACCAGACAAAAAA
CATCAGAAAGAACCTCCATTCTTggatgggatatGAACCCATCCTGATAATGGACA
GTACAGCCTATAATGCTGCCAGAAAAGACAGCTGGACTGTCAATGACATACAGAAGTTA
GTGGGAAATTGAATTGGCAAGTCAGATTATGCAGGGATTAAAGTAAAGCAGTTATGT
AAACTCCTTAGAGGAACCAAAGCACTAACAGAAGTAATACCAACTAACAGAAGCAGAG
CTAGAACTGGCAGAAAACAGGGAGATTCTAAAAGAACAGTACATGAAGTATATTGAC
CCATCAAAGACTTAGTAGCAGAAATACAGAAGCAGGGCAAGGCCATGGACATATCAA
ATTATCAAGAGCCATTAAAAATCTGAAAACAGGAAAGTATGCAAGGATGAGGGTGCC
CACACTAATGATGTAACAGAGGCAGTGCAAAAAGTATCCACAGAAAGCATA
GTAATATGGGAAAGATTCTAAATTAACTACCCATAACAAAGGAAACATGGGAAGCA
TGGTGGATGGAGTATTGGCAAGCTACCTGGATTCTGAGTGGGAGTTGTCAATACCCCT
CCCTTAGTGAATTATGGTACCAAGTTAGAGAAACCCATAGTAGGAGCAGAAACTTTC
TATGTAGATGGGCAGCTAATAGGGAGACTAAATTAGGAAAAGCAGGATATGTTACTGAC
AGAGGAAGACAAAAGTTGTCTCATAGCTGACACAACAAATCAGAAGACTGAATTACAA
GCAATTCTAGCTTGCAGGATTGGGATTAGAAGTAAACATAGTAACAGACTCACAA
TATGCATTAGGAATCATTCAAGCACAACCAAGATAAGAGTGAATCAGAGTTAGTCAGTCAA
ATAATAGAGCAGTTAATAAAAAGGAAAAGGTCTACCTGGCATGGTACCCAGCACACAAA
GGAATTGGAGGAAATGAACAAGTAGATAATTAGTCAGTGCTGGAATCAGGAAAGTACTA

FIG. 68

(SEQ ID NO:77)

GagProtMod_SF2 (GP1)

GTCGACGCCACCATTGGCGCCCGCGCCAGCGTGCTGAGCGGCGGCAGCTGGACAAGTGG
 GAGAAGATCCGCCTGCGCCCCGGCGCAAGAAGAAGTACAAGCTGAAGCACATCGTGTGG
 GCCAGCCGCGAGCTGGAGCGCTTCGCCGTGAACCCCGCCTGCTGGAGACCAGCGAGGGC
 TGCCGCCAGATCCTGGGCCAGCTGCAGCCCAGCCTGCAGACCGGCAGCGAGGAGCTGC
 AGCCTGTACAACACCGTGGCCACCTGTACTGCGTGCACCAGCGCATCGACGTCAAGGAC
 ACCAAGGAGGCCCTGGAGAAGATCGAGGAGGAGCAGAACAGTCCAAGAAGAAGGCCAG
 CAGGCCGCCGCCGCCGGCACCGCAACAGCAGCCAGGTGAGCCAGAACTACCCCATC
 GTGCAGAACCTGCAGGCCAGATGGTGCACCAGGCCATCAGCCCCGCAACCTGAACGCC
 TGGGTGAAGGTGGTGGAGGAGAAGGCCCTCAGCCCCGAGGTGATCCCCATGTTAGCGCC
 CTGAGCGAGGCCACCCCCCAGGACCTGAACACGATGTTAACACCGTGGCGGCCAC
 CAGGCCGCATGCAGATGCTGAAGGAGACCATCAACGAGGAGGCCGAGTGGGACCGC
 GTGCACCCCGTGCACGCCGCCCATGCCCGGCCAGATGCGCGAGCCCCGCCAGC
 GACATGCCGCCACCACAGCACCTGCAGGAGCAGATGGCTGGATGACCAACAACCC
 CCCATCCCCGTGGCGAGATCTACAAGACCCCTGCGCGCTGAGCAGGCCAGGACGTGAAG
 CGGATGTACAGCCCCACCAAGCATTGGACATCCGCAGGGCCCCAAGGAGGCCCTCCGC
 GACTACGTGGACCGCTTCTACAAGACCCCTGCGCGCTGAGCAGGCCAGGACGTGAAG
 AACTGGATGACCGAGACCCCTGCTGGTGCAGAACGCCAACCCGACTGCAAGACCATTG
 AAGGCTCTGGCCCCGGCCACCTGGAGGAGATGATGACGCCCTGCCAGGGCGTGGC
 GGCCCCGGCCACAAGGCCCGTGTGGCCAGGCGATGAGCCAGGTGACGAACCCGGCG
 ACCATCATGATGCAGCGCGCAACTCCGCAACCAGCGGAAGACCGTCAAGTGTCAAC
 TGCGGCAAGGAGGCCACCCGCAGGAACCTGCCGCCCCCGCAAGAAGGGCTGCTGG
 CGCTGCCGCCGAAGGACACCAAATGAAAGATTGACTGAGAGACAGGCTAATTTTA
 GGGAAAGATCTGGCCTTCTACAAGGAAAGGCCAGGGATTTCAGAGCAGACCAGAG
 CCAACAGCCCCACCAGAACAGAGAGCTTCAGGTTGGGAGGAGAAAACAACCTCCCTCAG
 AAGCAGGAGCCGATAGACAAGGAACGTATCCTTAACTCCCTCAGATCACTCTTGGC
 AACGACCCCTCGTCACAGTAAGGATCGCGGCCAGCTCAAGGAGGCGCTGCTGACACCG
 GCGCCGACGACCCGTGCTGGAGGAGATGAAACCTGCCCGCAAGTGGAAAGGCCAAGATGA
 TCGCGGGATCGGGGCTTCATCAAGGTGCGGCAGTACGACCAAGATCCCCGTGGAGATCT
 GCGGCCACAAGGCCATGGCACCGTGTGGAGGCCCCACCCCGTGAACATCATCGGCC
 GCAACCTGCTGACCCAGATCGGCTGCACCCCTGAACCTCCCCATCAGCCCCATCGAGACGG
 TGCCCGTGAAGCTGAAGCCGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCTGTAAG
 AATTC

FIG. 69

(SEQ ID NO:78)

GagProtMod_SF2 (GP2)

GTCGACGCCACCATTGGGCGCCCGGCCAGCGTGCTGAGCGGCGCGAGCTGGACAAGTGG
 GAGAAGATCCGCCTGCGCCCCGGCGGCAAGAAGAAGTACAAGCTGAAGCACATCGTGTGG
 GCCAGCCGCGAGCTGGAGCGCTTCGCCGTGAACCCCGGCTGCTGGAGACCGAGCGAGGGC
 TGCCGCCAGATCCTGGGCCAGCTGCAGCCCAGCCTGCAGACCGGCAGCGAGGAGCTGCGC
 AGCCTGTACAACACCCTGGCCACCCCTGTACTGCGTGCACCAGCGCATCGACGTCAAGGAC
 ACCAAGGAGGCCCTGGAGAAGATCGAGGAGGAGCAGAACAGTCCAAGAAGAAGGCCAG
 CAGGCCGCCGCCGCCGGCACCGGCAACAGCAGCCAGGTGAGCCAGAACTACCCCATC
 GTGAGAACCTGCAGGGCCAGATGGTGACCCAGGCCATCAGCCCCGCACCCCTAACGCC
 TGGGTGAAGGTGGTGGAGGAGAAGGCCCTCAGCCCCGAGGTGATCCCCATGTTCAGGCC
 CTGAGCGAGGGGCCACCCCCCAGGACCTGAACACGATGTTAACACCCGTGGCCAGGCCAC
 CAGGCCGCATGCAGATGCTGAAGGAGACCATCAACGAGGAGGCCAGTGGCAGGCCAGC
 GTGCACCCCGTGCACGCCGCCCATGCCCGGCCAGATGCGCAGGCCAGGCCAGC
 GACATGCCGGCACCAACCAGCACCTGCAGGAGCAGATGGCTGGATGACCAACAACCC
 CCCATCCCCGTGGCGAGATCTACAAGCGGTGGATCATCCTGGCCTGAACAAAGATGTG
 CGGATGTACAGCCCCACCAGCATCCTGGACATCGCCAGGGCCCCAAGGAGGCCCTCCGC
 GACTACGTGGACCGCTTCTACAAGACCCCTGCGCGCTGAGCAGGCCAGCCAGGACGTGAAG
 AACTGGATGACCGAGACCCCTGCTGGTGCAGAACGCCAACCCCCGACTGCAAGACCATCCTG
 AAGGCTCTCGGCCCCCGGGCCACCCCTGGAGGAGATGATGACCGCCTGCCAGGGCGTGGC
 GGCCCCGGCCACAAGGCCCGTGTGGCCAGGGATGAGCCAGGTGACGAACCCGGCG
 ACCATCATGATGCAGCGCGCAACTTCCGCAACCAGCGGAAGACCGTCAAGTGTCAAC
 TGCGGCAAGGAGGGCACACCGCCAGGAACCGCCAGGAACTGCCGCCGCCCCCGCAAGAAGGGCTGCTGG
 CGCTCGGCCCGGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTA
 GGGAAAGATCTGGCCTTCTACAAGGAAAGGCCAGGGAAATTTCTTCAGAGCAGACCAGAG
 CCAACAGCCCCACCAGAAGAGAGCTTCAGGTTGGGAGGAGAAAACAACCTCCCTCTCAG
 AAGCAGGAGCCGATAGACAAGGAACGTATCCTTAACCTCCCTCAGATCACTCTTGGC
 AACGACCCCTCGTCACAGTAAGGATCGGGGGCAACTCAAGGAAGCGCTGCTCGATAACAG
 GAGCAGATGATACAGTATTAGAAGAAATGAATTGCCAGGAAAATGAAACCAAAATGA
 TAGGGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATAACCTGTAGAAATCT
 GTGGACATAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAA
 GAAATCTGTTGACCCAGATCGGCTGCACCTTGAACCTCCCCATCAGCCCTATTGAGACGG
 TGCCCGTGAAGTTGAAGCCGGGATGGACGGCCCCAAGGTCAAGCAATGGCATTGTAAG
 AATTC

FIG. 70

(SEQ ID NO:79)

FS(+)_ProtInact_RTopt_YM

GCGGCCGCGAAGGACACAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTAGGGA
 AGATCTGGCCTTCCTACAAGGGAGGCCAGGGATTTCAGAGCAGACCAGAGCCAA
 CAGCCCCACCAGAAGAGAGCTTCAGGTTGGGAGGAGAAAACAACCTCCCTCAGAAC
 AGGAGCCGATAGACAAGGAACGTATCCTTAACCTCCCTCAGATCACTCTTGGCAACG
 ACCCTCGTCACAATAAGGATCGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGC
 AGATGATACAGTATTAGAAGAAATGAATTGCCAGGAAAATGAAACCAAAATGATAGG
 GGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCTGTGG
 ACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGAAAGAAA
 TCTGTTGACCCAGATCGGCTGCACCTTGAACCTCCCACATCAGCCCTATTGAGACGGTGCC
 CGTGAAGTTGAAGCCGGGATGGACGGCCCCAAGGTCAAGCAATGCCATTGACCGAGGA
 GAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAA
 GATCGGCCCCGAGAACCCCTACAACACCCCCGTGTCGCCATCAAGAAGAAGGACAGCAC
 CAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGA
 GGTGCAGCTGGCATCCCCACCCCGCCGCCTGAAGAAGAAGAGCGTGACCGTGCT
 GGACGTGGCGACGCCTACTTCAGCGTCCCCCTGGACAAGGACTTCCGCAAGTACACCGC
 CTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAAGTACAACGTGCT
 GCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTCCAGAGCAGCATGACCAAGATCCTGGA
 GCCCTTCCGCAAGCAGAACCCCGACATCGTATCTACCGAGCCCCCTGTACGTGGCAG
 CGACCTGGAGATCGGCCAGCACCGACCAAGATCGAGGAGCTGCCAGCACCTGCTGCC
 CTGGGGCTTCACCACCCCGACAAGAAGCACCAGAAGGAGCCCCCTTCTGTGGATGGG
 CTACGAGCTGCACCCGACAAGTGGACCGTGAGCCATCATGCTGCCAGAAGGACAG
 CTGGACCGTGAAACGACATCCAGAAGCTGGTGGCAAGCTGAACCTGGCCAGCCAGATCTA
 CGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCCGGCACCAAGGCCCTGACCGA
 GGTGATCCCCCTGACCGAGGGCCAGCTGGAGCTGGCCAGAACCGCGAGATCCTGAA
 GGAGCCCGTGACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCAGATCCAGAA
 GCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAAGGAGCCCTCAAGAACCTGAAGAC
 CGGCAAGTACGCCCGCATGCCGGCGCCACACCAACGACGTGAAGCAGCTGACCGAGGC
 CGTGCAGAAGGTGAGCACCGAGAGCATCGTATCTGGGCAAGATCCCCAAGTCAAGCT

FIG. 71A

(SEQ ID NO:80)

116 / 131

GCCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGAT
CCCCGAGTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAA
GGAGGCCATCGTGGCGCCGAGACCTTCTACGTGGACGGCGCCCAACCGCGAGACCAA
GCTGGGCAAGGCCGGTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATGCCGA
CACCAACCAACCAGAACGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCT
GGAGGTGAACATCGTACCGACAGCCAGTACGCCCTGGCATCATCCAGGCCAGCCCAG
CAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAACGGT
GTACCTGGCCTGGGTGCCGCCACAAGGGCATGGCGAACGAGCAGGTGGACAAGCT
GGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCCTGAACGGCATCGATGGCGGCATCGTGA
CTACCAGTACATGGACGACCTGTACGTGGCAGCGGCCCTAGGATCGATTAAAAGCT
TCCCGGGGCTAGCACCGGTGAATT

FIG. 71B
(SEQ ID NO:80)

FS(+)_ProtInact_RTopt_YMWM

GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTAGGGA
 AGATCTGGCCTCCTACAAGGGAAAGGCCAGGGATTTCAGAGCAGACCAGAGCCAA
 CAGCCCCACCAGAAGAGAGCTTCAGGTTGGGAGGAGAAAACAACCTCCCTCAGAACG
 AGGAGCCGATAGACAAGGAACTGTATCCTTAACCTCCCTCAGATCACTCTTGGCAACG
 ACCCCTCGTCACAATAAGGATCGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGC
 AGATGATACTAGTATTAGAAGAAATGAATTGCCAGGAAATGAAACCAAAATGATAGG
 GGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACTGTAGAAATCTGTGG
 ACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAA
 TCTGTTGACCCAGATCGGCTGCACCTTGAACCTCCCACATCAGCCCTATTGAGACGGTGCC
 CGTGAAGTTGAAGCCGGGATGGACGGCCCCAAGGTCAAGCAATGCCATTGACCGAGGA
 GAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAA
 GATCGGCCCCGAGAACCCCTACAAACACCCCCGTGTCGCCATCAAGAAGAAGGACAGCAC
 CAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGA
 GGTGCAGCTGGCATCCCCACCCCGCCGGCTGAAGAAGAAGAAGACGCGTACCGTGCT
 GGACGTGGCGACGCCTACTTCAGCGTCCCCCTGGACAAGGACTTCCGCAAGTACACCGC
 CTTCACCATCCCCAGCATCAACAAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCT
 GCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTCCAGAGCAGCATGACCAAGATCCTGGGA
 GCCCTCCGCAAGCAGAACCCGACATCGTGTACCGAGGCCCCCTGTACGTGGCAG
 CGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCG
 CTGGGGCTTCACCACCCCGACAAGAACGACCAAGAGGAGCCCCCTTCTGCCATCGA
 GCTGCACCCGACAAGTGGACCGTGCAGCCATCATGCTGCCGAGAACGGACAGCTGGAC
 CGTGAACGACATCCAGAAGCTGGTGGCAAGCTGAACCTGGCCAGCCAGATCTACGCCGG
 CATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCCGACCAAGGCCCTGACCGAGGTGAT
 CCCCCCTGACCGAGGAGGCCAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGGAGCC
 CGTGCACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCGAGATCCAGAACGG
 CCAGGGCCAGTGGACCTACCAGATCTACCAAGGAGCCCTCAAGAACCTGAAGACCGGCAA
 GTACGCCCGCATGCGCGGCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCA
 GAAGGTGAGCACCGAGAGCAGTCGTGATCTGGGCAAGATCCCCAAGTTCAAGCTGCCAT

FIG. 72A

(SEQ ID NO:81)

CCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCGA
GTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCC
CATCGTGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGG
CAAGGCCGGCTACGTGACCGACCAGGGCCGGCAGAAGGTGGTGAGCATTGGCGACACCAC
CAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGT
GAACATCGTGACCGACAGCCAGTACGCCCTGGCATCATCCAGGCCAGCCGACAAGAG
CGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCT
GGCCTGGGTGCCGCCACAAGGGCATGGCGAACGAGCAGGTGGACAAGCTGGTGAG
CGCCGGCATCCGCAAGGTGCTGTTCTGAACGGCATCGATGGCGGCATCGTATCTACCA
GTACATGGACGACCTGTACGTGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGG
GGCTAGCACCGGTGAATT

FIG. 72B
(SEQ ID NO:81)

FS(-)_ProtMod_RTopt_YM

GC GGCC CGA AGGAC ACCAAAT GAAAG ATT GCA CTGAGAGACAGGCTAATTCTTCCGCG
AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGAGTT CAGCAGCGAGCACCCGCGCCA
ACAGCCCCACCCGCCGAGCTCAGGTGTGGGCAGCGAGAACAAACAGCCTGAGCGAGG
CCGGCGCCGACCGCCAGGGCACCGTGAGCTCAACTCCCCAGATCACCTGTGGCAGC
GCCCCCTGGT GACC ATCAGGATCGCGGCCAGCTCAAGGAGGCCTGCTCGACACCGCG
CCGACGACACCGTGCTGGAGGAGATGAACCTGCCGGCAAGTGGAGGCCAAGATGATCG
GCGGGATCGGGGCTTCATCAAGGTGCGGAGTACGACCAGATCCCCGTGGAGATCTGCG
GCCACAAGGCCATCGGCACCGTGCTGGTGGCCCCACCCCGTGAACATCATCGGCCGCA
ACCTGCTGACCCAGATCGGCTGCACCTGAACCTCCCCATCAGCCCCATCGAGACGGTGC
CCGTGAAGCTGAAGCCGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCTGACCGAGG
AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA
AGATCGGCCCGAGAACCCCTACAACACCCCCGTGTCGCCATCAAGAAGAAGGACAGCA
CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTCTGG
AGGTGCAGCTGGCATCCCCACCCGCCGCTGAAGAAGAAGAGCGTGACCGTGC
TGGACGTGGCGACGCCTACTTCAGCGTCCCCCTGGACAAGGACTTCGCAAGTACACCG
CCTTCACCATCCCCAGCATCAACAAGAGACCCCCGGCATCCGCTACCAGTACAACGTGC
TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTCCAGAGCAGCATGACCAAGATCTGG
AGCCCTTCCGCAAGCAGAACCCCGACATCGTGA TCTACCAGGCCCCCTGTACGTGGCA
GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC
GCTGGGGCTTCACCACCCCGACAAGAACGACCAGAAGGAGCCCCCTTCCTGTGGATGG
GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCGAGAACGGACA
GCTGGACCGTGAACGACATCCAGAAGCTGGTGGCAAGCTGAACCTGGCCAGCCAGATCT
ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCCAGCACCAAGGCCCTGACCG
AGGTGATCCCCCTGACCGAGGAGGCCAGTGGACCTACCAAGATCTACCAGGAGCCCTTAAGAACCTGAAGA
AGGAGCCCGTGCACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCAGATCCAGA
AGCAGGGCCAGGCCAGTGGACCTACCAAGATCTACCAGGAGCCCTTAAGAACCTGAAGA
CCGGCAAGTACGCCGCATGCGCGGCCACACCAACGACGTGAAGCAGCTGACCGAGG
CCGTGCAGAAGGTGAGCACCAGAGCATCGTGA TCTGGGCAAGATCCCCAAGTCAAGC

FIG. 73A
(SEQ ID NO:82)

120 / 131

TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA
TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA
AGGAGCCCATCGTGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA
AGCTGGCAAGGCCGGTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATGCCG
ACACCAACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGCCCTGCAGGACAGCGGCC
TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGCATCATCCAGGCCAGCCCG
ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG
TGTACCTGGCCTGGGTGCCGCCACAAGGGCATGGCGAACGAGCAGGTGGACAAGC
TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCTGAACGGCATCGATGGCGGCATGTGA
TCTACCAAGTACATGGACGACCTGTACGTGGCAGCGGCCCTAGGATCGATTAAAAGC
TTCCCGGGCTAGCACCGGTGAATT

FIG. 73B
(SEQ ID NO:82)

FS(-)_ProtMod_RTopt_YMWM

GCGGCCGCGAAGGACACAAATGAAAGATTGCACTGAGAGACAGGCTAATTCTCCGCG
 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGAGTTCAGCAGCAGCAGACCCGCGCCA
 ACAGCCCCACCCGCCGAGCTGCAGGTGTGGGGCGAGAACAACAGCCTGAGCGAGG
 CGCGGCCGACCGCCAGGGCACCGTGAGCTCAACTCCCCAGATCACCTGTGGCAGC
 GCCCCCTGGTGACCATCAGGATCGCGGCCAGCTCAAGGAGGCCTGCTCGACACCGCG
 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCGCAAGTGGAGGCCAAGATGATCG
 CGGGATCGGGGCTTCATCAAGGTGCGGAGCTACGACCAGATCCCCGTGGAGATCTGCG
 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCACCCCCGTGAACATCATCGGCCGCA
 ACCTGCTGACCCAGATCGGCTGCACCCCTGAACCTCCCCATCAGCCCCATCGAGACGGTGC
 CCGTGAAGCTGAAGCCGGGATGGACGCCCAAGGTCAAGCAGTGGCCCTGACCGAGG
 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA
 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTCGCCATCAAGAAGAAGGACAGCA
 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGG
 AGGTGCAGCTGGCATCCCCACCCGCCGCTGAAGAAGAAGAGCGTGACCGTGC
 TGGACGTGGCGACGCCTACTTCAGCGTCCCCCTGGACAAGGACTTCCGCAAGTACACCG
 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCGCTACCAAGTACAACGTGC
 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATTTCCAGAGCAGCATGACCAAGATCCTGG
 AGCCCTTCCGCAAGCAGAACCCGACATCGTGTACTACCAAGGCCCCCTGTACGTGGCA
 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCCAGCACCTGCTGC
 GCTGGGCTTACCAACCCCCGACAAGAACGACCAAGAAGGAGCCCCCTTCCTGCCATCG
 AGCTGCACCCGACAAGTGGACCGTGCAGCCATCATGCTGCCGAGAAGGACAGCTGGA
 CCGTGAACGACATCCAGAACGACTGGTGGCAAGCTGAACCTGGCCAGCCAGATCTACGCCG
 GCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCCGGCACCAAGGCCCTGACCGAGGTGA
 TCCCCCTGACCGAGGAGGCCAGCTGGAGCTGGCCAGAACCGCGAGATCCTGAAGGAGC
 CCGTGCACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCAGATCCAGAACGAGG
 GCCAGGGCCAGTGGACCTACCAGATCTACCAAGGAGCCCTCAAGAACCTGAAGACCGGCA
 AGTACGCCGCATGCCGGCGCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC
 AGAAGGTGAGCAGCAGCATCGTGTACTGGGCAAGATCCCCAAGTCAAGCTGCCA

FIG. 74A

(SEQ ID NO:83)

122 / 131

TCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCG
AGTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC
CCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGG
GCAAGGCCGGCTACGTGACCGACCGACGGGGCCGGCAGAAGGTGGTGAGCATGCCGACACCA
CCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGG
TGAACATCGTGACCGACAGCCAGTACGCCCTGGCATCATCCAGGCCAGCCCACAAGA
GCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC
TGGCCTGGGTGCCCGCCCACAAGGGCATGGCGAACGAGCAGGTGGACAAGCTGGTGA
GCGCCGGCATCCGCAAGGTGCTGTTCTGAACGGCATCGATGGCGGCATCGTATCTACC
AGTACATGGACGACCTGTACGTGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG
GGGCTAGCACCGGTGAATTC

FIG. 74B

(SEQ ID NO:83)

123 / 131

FS(-)_ProtMod_RTopt(+)

GC GGCCCGAAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTCTTCCGCG
AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCAGTCAGCAGCAGCAGACCCGCGCCA
ACAGCCCCACCCGCCGAGCTGCAGGTGTGGGCGGCGAGAACAAACAGCCTGAGCGAGG
CCGGCGCCGACCGCCAGGGCACCGTGAGCTCAACTCCCCAGATCACCTGTGGCAGC
GCCCTGGTGACCATCAGGATCGCGGCCAGCTCAAGGAGCGCTGCTGACACCGCG
CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGCAAGTGGAGCCAAAGATGATCG
GCGGGATCGGGGCTTCATCAAGGTGGCGAGTACGACCAGATCCCCGTGGAGATCTGCG
GCCACAAGGCCATCGGCACCGTGCTGGTGGCCCCACCCCCGTGAACATCATCGGCCGCA
ACCTGCTGACCCAGATCGGCTGCACCCATCACGCCCCATCGAGACGGTGC
CCGTGAAGCTGAAGCCGGGATGGACGGCCCAAGGTCAAGCAGTGGCCCTGACCGAGG
AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA
AGATCGGCCCCAGAACCCCTACAACACCCCCGTGTTGCCATCAAGAAGAAGGACAGCA
CCAAGTGGCGCAAGCTGGACTTCCGCAGCTGAACAAGCGCACCCAGGACTTCTGGG
AGGTGCAGCTGGCATCCCCACCCGCCGCTGAAGAAGAAGAGCGTGAACCGTGC
TGGACGTGGCGACGCCTACTTCAGCGTCCCCGGACAAGGACTTCCGCAAGTACACCG
CCTTCACCATCCCCAGCATCAACAACGAGACCCCGGCATCCGCTACCAGTACAACGTGC
TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTCCAGAGCAGCATGACCAAGATCCTGG
AGCCCTTCCGCAAGCAGAACCCGACATCGTGTACTACCGAGTACATGGACGACCTGTACG
TGGCGACCTGGAGATCGGCCAGCACCGACCAAGATCGAGGAGCTGCGCCAGCACC
TGCTGCGCTGGGCTTCACCACCCCGACAAGAAGCACCAGAAGGAGCCCCCTTCCTGT
GGATGGCTACGAGCTGCACCCGACAAGTGGACCGTGCAGCCATCATGCTGCCGAGA
AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGCAAGCTGAACCTGGCCAGCC
AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCCGAGCACCAAGGCC
TGACCGAGGTGATCCCCCTGACCGAGGAGGCCAGTGGACCTACCAGATCTACCGAGGCC
TCCTGAAGGAGCCCCTGCACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCGAGA
TCCAGAAGCAGGCCAGGGCCAGTGGACCTACCAGATCTACCGAGGCC
TGAAGACCGGCAAGTACGCCGCATGCGCGGCCACACCAACGACGTGAAGCAGCTGA
CCGAGGCCGTGCAAGGTGAGCAGCAGAGAGCATCGTGTACTGGGCAAGATCCCCAAGT
TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCCTGGTGGAGTGGAGTACTGGCAGGCC
CCTGGATCCCCAGTGGAGTTCGTGAACACCCCCCTGGTGAAGCTGTGGTACCGAGC
TGGAGAAGGAGCCATCGTGGCGCCGAGACCTCTACGTGGACGGCGCCAAACCGCG

FIG. 75A

(SEQ ID NO:84)

124 / 131

AGACCAAGCTGGCAAGGCCGGTACGTGACCGACCGGGCCGGCAGAAGGTGGTGAGCA
TCGCCGACACCACCAACCAGAACGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA
GCGGCCTGGAGGTGAACATCGTACCGACAGCCAGTACGCCCTGGCATCATCCAGGCC
AGCCCGACAAGAGCGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG
AGAAGGTGTACCTGGCCTGGGTGCCGCCACAAGGGCATCGCGGCAACGAGCAGGTGG
ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCTGAACGGCATCGATGGCGGCA
TCGTGATCTACCAAGTACATGGACGACCTGTACGTGGCAGCGGCGGCCCTAGGATCGATT
AAAAGCTTCCCAGGGCTAGCACCGGTGAATTC

FIG. 75B
(SEQ ID NO:84)

125 / 131

Tat_wt_SF162 (wildtype)

ATGGAGCCAGTAGATCCTAGATTAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTAAGA
 CTGCTTGACAAATTGCTATTGTAAGGTGCTTCATTGCCAAGTTGTTCTACAAAC
 AAAAGGCTTAGGCATCTCTATGGCAGGAAGAACGGAGACAGCGACGAAGAGCTCCT
 CCAGACAGTGGAGTTCATCAAGTTCTACCAAAGCAACCCGCTTCCCAGCCCCAAGG
 GGACCCGACAGGCCGAAGGAATCGAAGAAGAAGGTGGAGAGAGACAGAGACAGA
 TCCAGTCCATTAG

FIG. 76
 (SEQ ID NO:85)

Tat_SF162

MEPVDPRLPWKHPGSQPKACTNCYCKKCFHCQVCFITKGLGISYGRKKRRQRRAPPDSE
 VHQVSLPKQPASQPQGDPTGPKESKKVERETETDPVH

FIG. 77
 (SEQ ID NO:86)

Tat_SF162_opt

ATGGAGCCCGTGGACCCCCGCTGGAGCCCTGGAAGCACCCCGCAGCCAGCCAAAGAC
 CGCCTGCACCAACTGCTACTGCAAGAAGTGTGCTTCCACTGCCAGGTGTGCTTCATCACCA
 AAGGGCCTGGGCATCAGCTACGGCCGCAAGAACGCGCCGCCAGCGCCGCCGCGCCCCCCCC
 CGACAGCGAGGTGCACCAGGTGAGCCTGCCAAGCAGCCGCCAGCCAGCCCCAGGGCG
 ACCCCACCGGCCCAAGGAGAGCAAGAAGAACGGTGGAGCGCGAGACCGAGACCGACCCCC
 GTGCACTAG

FIG. 78
 (SEQ ID NO:87)

Tat_Cys22_SF162_opt

ATGGAGCCCGTGGACCCCCGCTGGAGCCCTGGAAGCACCCCGCAGCCAGCCAAAGAC
 CGCCgGCACCAACTGCTACTGCAAGAAGTGTGCTTCCACTGCCAGGTGTGCTTCATCACCA
 AGGGCCTGGGCATCAGCTACGGCCGCAAGAACGCGCCGCCAGCGCCGCCGCGCCCCCCCC
 GACAGCGAGGTGCACCAGGTGAGCCTGCCAAGCAGCCGCCAGCCAGCCCCAGGGCGA
 CCCACCGGCCCAAGGAGAGCAAGAACGGTGGAGCGCGAGACCGAGACCGACCCCCG
 TGCACTAG

FIG. 79
 (SEQ ID NO:88)

126 / 131

Alignment GagMod vs GP1 GP2

GagMod. SF2	(1)	1	10	20	30	40	50	60	76
	(1)	ATGGGCCCGGGCCAGGGCTGAGGGGGAGCTGGACAAGTGGAGAAGATA	TCGGCTGCGCCCCGGGGCA						Section 1
GagProtMod. SF2 (GP1)	(1)	ATGGGCGCCCGGGCCAGGGCTGAGGGGGAGCTGGACAAGTGGAGAAGAT	TCGGCTGCGCCCCGGGGCA						
GagProtMod. SF2 (GP2)	(1)	ATGGGGCGCCGGCCAGGGCTGAGGGGGAGCTGGACAAGTGGAGAAGAT	TCGGCTGCGCCCCGGGGCA						
Consensus	(1)	ATGGGGCGCCGGCCAGGGCTGAGGGGGAGCTGGACAAGTGGAGAAGAT	TCGGCTGCGCCCCGGGGCA						
									Section 2
GagMod. SF2	(77)	77	90	100	110	120	130	140	152
	(77)	AGAAGAAGTACAAAGCTGAAGGCACATCGTGTGGGGCAGGCCGGAGCT	TGCCGTGAAACCCCCGGCTGCT						
GagProtMod. SF2 (GP1)	(77)	AGAAGAAGTACAAAGCTGAAGGCACATCGTGTGGGGCAGGCCGGAGCT	TGCCGTGAAACCCCCGGCTGCT						
GagProtMod. SF2 (GP2)	(77)	AGAAGAAGTACAAAGCTGAAGGCACATCGTGTGGGGCAGGCCGGAGCT	TGCCGTGAAACCCCCGGCTGCT						
Consensus	(77)	AGAAGAAGTACAAAGCTGAAGGCACATCGTGTGGGGCAGGCCGGAGCT	TGCCGTGAAACCCCCGGCTGCT						
									Section 3
GagMod. SF2	(153)	153	160	170	180	190	200	210	228
	(153)	GGAGACCAGGGAGGGCTGCCAGATCCTGGGCCAGCTGCAGCCCAGC	GGAGCTGCAGGGAGCTGGCG						
GagProtMod. SF2 (GP1)	(153)	GGAGACCAGGGAGGGCTGCCAGATCCTGGGCCAGCTGCAGCCCAGC	GGAGCTGCAGGGAGCTGGCG						
GagProtMod. SF2 (GP2)	(153)	GGAGACCAGGGAGGGCTGCCAGATCCTGGGCCAGCTGCAGCCCAGC	GGAGCTGCAGGGAGCTGGCG						
Consensus	(153)	GGAGACCAGGGAGGGCTGCCAGATCCTGGGCCAGCTGCAGCCCAGC	GGAGCTGCAGGGAGCTGGCG						
									Section 4
GagMod. SF2	(229)	229	240	250	260	270	280	290	304
	(229)	AGCCGTGTACAACACCGTGGCCACCTGTACTGCGTCACGTCAAGG	ACACCAAGGAGGCCCTGG						
GagProtMod. SF2 (GP1)	(229)	AGCCGTGTACAACACCGTGGCCACCTGTACTGCGTCACGTCAAGG	ACACCAAGGAGGCCCTGG						
GagProtMod. SF2 (GP2)	(229)	AGCCGTGTACAACACCGTGGCCACCTGTACTGCGTCACGTCAAGG	ACACCAAGGAGGCCCTGG						
Consensus	(229)	AGCCGTGTACAACACCGTGGCCACCTGTACTGCGTCACGTCAAGG	ACACCAAGGAGGCCCTGG						
									Section 5
GagMod. SF2	(305)	305	310	320	330	340	350	360	380
	(305)	AGAAGATCGAGGAGGGAGCAGAACAGTCCAAGAACAGTCCAGG	GGCCGGCCAGGAGGCCACGG						
GagProtMod. SF2 (GP1)	(305)	AGAAGATCGAGGAGGGAGCAGAACAGTCCAAGAACAGTCCAGG	GGCCGGCCAGGAGGCCACGG						
GagProtMod. SF2 (GP2)	(305)	AGAAGATCGAGGAGGGAGCAGAACAGTCCAAGAACAGTCCAGG	GGCCGGCCAGGAGGCCACGG						
Consensus	(305)	AGAAGATCGAGGAGGGAGCAGAACAGTCCAAGAACAGTCCAGG	GGCCGGCCAGGAGGCCACGG						

FIG. 80A

127 / 131

Alignment GagMod vs GP1 GP2

SUBSTITUTE SHEET (RULE 26)

FIG. 80B

128 / 131

Alignment GagMod vs GP1_GP2

										Section 11
(761)	761	770	780	790	800	810	820	830	836	
GagMod.	SF2	(761)	ACAACCCCCCATCCCCGTGGGAGATCTACAAGGGGGATCATCCTGGGCTGAACAAAGATCGTGGGATGTA							
GagProtMod.	SF2 (GP1)	(761)	ACAACCCCCCATCCCCGTGGGAGATCTACAAGGGGGATCATCCTGGGCTGAACAAAGATCGTGGGATGTA							
GagProtMod.	SF2 (GP2)	(761)	ACAACCCCCCATCCCCGTGGGAGATCTACAAGGGGGATCATCCTGGGCTGAACAAAGATCGTGGGATGTA							
Consensus		(761)	ACAACCCCCCATCCCCGTGGGAGATCTACAAGGGGGATCATCCTGGGCTGAACAAAGATCGTGGGATGTA							Section 12
(837)	837	850	860	870	880	890	900	910	912	
GagMod.	SF2	(837)	CAGCCCCACCCAGCATCTGGACATCCGCCAGGGCCCAGGAGCCCAGGGCCCTTCGGGACTACGTGGACGCC							
GagProtMod.	SF2 (GP1)	(837)	CAGCCCCACCCAGCATCTGGACATCCGCCAGGGCCCAGGAGCCCAGGGCCCTTCGGGACTACGTGGACGCC							
GagProtMod.	SF2 (GP2)	(837)	CAGCCCCACCCAGCATCTGGACATCCGCCAGGGCCCAGGAGCCCAGGGCCCTTCGGGACTACGTGGACGCC							
Consensus		(837)	CAGCCCCACCCAGCATCTGGACATCCGCCAGGGCCCAGGAGCCCAGGGCCCTTCGGGACTACGTGGACGCC							Section 13
(913)	913	920	930	940	950	960	970	980	988	
GagMod.	SF2	(913)	ACCCCTGCGCGCTGAGCAGGCCAGGCCAGGTGAAGAFAACTGGATGACCGAGAACCCCTGCGGTGCA							
GagProtMod.	SF2 (GP1)	(913)	ACCCCTGCGCGCTGAGCAGGCCAGGCCAGGTGAAGAFAACTGGATGACCGAGAACCCCTGCGGTGCA							
GagProtMod.	SF2 (GP2)	(913)	ACCCCTGCGCGCTGAGCAGGCCAGGCCAGGTGAAGAFAACTGGATGACCGAGAACCCCTGCGGTGCA							
Consensus		(913)	ACCCCTGCGCGCTGAGCAGGCCAGGCCAGGTGAAGAFAACTGGATGACCGAGAACCCCTGCGGTGCA							Section 14
(989)	989	1000	1010	1020	1030	1040	1050	1060		
GagMod.	SF2	(989)	CGGACTGCAAAGACCATTCTGAAGGCTCTGGGCCCCGGCCACCCCTGGAGGAGATGATGACCGCCTG							
GagProtMod.	SF2 (GP1)	(989)	CGGACTGCAAAGACCATTCTGAAGGCTCTGGGCCCCGGCCACCCCTGGAGGAGATGATGACCGCCTG							
GagProtMod.	SF2 (GP2)	(989)	CGGACTGCAAAGACCATTCTGAAGGCTCTGGGCCCCGGCCACCCCTGGAGGAGATGATGACCGCCTG							
Consensus		(989)	CGGACTGCAAAGACCATTCTGAAGGCTCTGGGCCCCGGCCACCCCTGGAGGAGATGATGACCGCCTG							Section 15
(1065)	1065	1070	1080	1090	1100	1110	1120	1130	1140	
GagMod.	SF2	(1065)	GGGGGGCCCCGGCCACAAAGGCCCGGTGGCTGGCCGAGGGCATGAGCCAGGTGACGAACCGGGATG							
GagProtMod.	SF2 (GP1)	(1065)	GGGGGGCCCCGGCCACAAAGGCCCGGTGGCTGGCCGAGGGCATGAGCCAGGTGACGAACCGGGATG							
GagProtMod.	SF2 (GP2)	(1065)	GGGGGGCCCCGGCCACAAAGGCCCGGTGGCTGGCCGAGGGCATGAGCCAGGTGACGAACCGGGATG							
Consensus		(1065)	GGGGGGCCCCGGCCACAAAGGCCCGGTGGCTGGCCGAGGGCATGAGCCAGGTGACGAACCGGGATG							

FIG. 80C

129 / 131

Alignment GagMod vs GP1_GP2

					Section 16
(1141)	1141	1150	1160	1170	1180
GagMod.	SF2 (1141)	CAGCGCGGCCAACTTCCGCAACCAGCGGAAGACCGGTCAAAGTGCTTCAC	TGCGGCCAACCGGCCACACGGCCAGGA	1190	1200
GagProtMod.	SF2 (GP1) (1141)	CAGCGCGGCCAACTTCCGCAACCAGCGGAAGACCGGTCAAAGTGCTTCAC	TGCGGCCAACCGGCCACACGGCCAGGA	1216	1216
GagProtMod.	SF2 (GP2) (1141)	CAGCGCGGCCAACTTCCGCAACCAGCGGAAGACCGGTCAAAGTGCTTCAC	TGCGGCCAACCGGCCACACGGCCAGGA		
Consensus	(1141)	CAGCGCGGCCAACTTCCGCAACCAGCGGAAGACCGGTCAAAGTGCTTCAC	TGCGGCCAACCGGCCACACGGCCAGGA		
					Section 17
(1217)	1217	1230	1240	1250	1260
GagMod.	SF2 (1217)	ACTGCCGCCCCCGCAAGAAGGGCTGCTGGCGCTGCAGGGCCACCAAGATG	AAGGGACTGCACCGAGCG	1270	1280
GagProtMod.	SF2 (GP1) (1217)	ACTGCCGCCCCCGCAAGAAGGGCTGCTGGCGCTGCAGGGCCACCAAA	TGAAAGATTGCACTGAGAG		
GagProtMod.	SF2 (GP2) (1217)	ACTGCCGCCCCCGCAAGAAGGGCTGCTGGCGCTGCAGGGCCACCAAA	TGAAAGATTGCACTGAGAG		
Consensus	(1217)	ACTGCCGCCCCCGCAAGAAGGGCTGCTGGCGCTGCAGGGCCACCAAA	TGAAAGATTGCACTGAGAG		
					Section 18
(1293)	1293	1300	1310	1320	1330
GagMod.	SF2 (1293)	CCAGGCCAACCTCCTGGCAAGATCTGGCCAGCTACAAGGGCCGGCAACT	TCCCTGCAGAGCCGCCAG	1340	1350
GagProtMod.	SF2 (GP1) (1293)	ACAGGGCTAAATTTTTAAAGGAAAGATCTGGCCCTTCCTACAAGGG	GAAGGCCAGGGAAATTTCCTCAAG		
GagProtMod.	SF2 (GP2) (1293)	ACAGGGCTAAATTTTTAAAGGAAAGATCTGGCCCTTCCTACAAGGG	GAAGGCCAGGGAAATTTCCTCAAG		
Consensus	(1293)	ACAGGGCTAAATTTTTAAAGGAAAGATCTGGCCCTTCCTACAAGGG	GAAGGCCAGGGAAATTTCCTCAAG		
					Section 19
(1369)	1369	1380	1390	1400	1410
GagMod.	SF2 (1369)	CCCACGGCCCCCGAGGAGACTTCGGCGAGCTTCAGGTTGGGGAGGA	AAAGGACCCAGGCCAGGAGCCATCG	1420	1430
GagProtMod.	SF2 (GP1) (1369)	CCAACAGCCCCCACAGAGAGACTTCAGGTTGGGGAGGA	AAAGGACCCCTCTCAAGGAGGCCGATAG		
GagProtMod.	SF2 (GP2) (1369)	CCAACAGCCCCCACAGAGAGACTTCAGGTTGGGGAGGA	AAAGGACCCCTCTCAAGGAGGCCGATAG		
Consensus	(1369)	CCAACAGCCCCCACAGAGAGACTTCAGGTTGGGGAGGA	AAAGGACCCCTCTCAAGGAGGCCGATAG		
					Section 20
(1445)	1445	1450	1460	1470	1480
GagMod.	SF2 (1445)	ACAAGGGAGCTGTACCCCTGACCAAGCCCTGTTCGGCACACGAC	CCCCCAGCCAGCAGTAA-----	1490	1500
GagProtMod.	SF2 (GP1) (1445)	ACAAGGGAACTGTATCCCTTAACCTCCCTCAAGTCACTTTGGCA	ACGACCCCTCGTCAGTAAGGATCGGGCC		
GagProtMod.	SF2 (GP2) (1445)	ACAAGGGAACTGTATCCCTTAACCTCCCTCAAGTCACTTTGGCA	ACGACCCCTCGTCAGTAAGGATCGGGGG		
Consensus	(1445)	ACAAGGGAACTGTATCCCTTAACCTCCCTCAAGTCACTTTGGCA	ACGACCCCTCGTCAGTAAGGATCGGGG		

FIG. 80D

130 / 131

Alignment GagMod vs GP1_GF2

Section 21									
(1521)	1521	1530	1540	1550	1560	1570	1580	1590	1596
GagMod_SF2(1510)	--								
GagProtMod_SF2(GP1)(1521)	CAGCTCAAGGAGGGCCTGCTGACACCGGCCGAGCACCGTGTGGAGGATGAACCTGCCGGCAAGTGGAA								
GagProtMod_SF2(GP2)(1521)	CAACTCAAGGAAGCGCTGCTGATACAGGAGCAGATGATACTAGTAAGAATGAAATGAAATGAAATGGAA								
Consensus(1521)	CACTCAAGGAAGCGCTGCTGAAACAGTACAGGAGCAGATGATACTAGTAAGAATGAAATGAAATGGAA	AC	GG	GC	GA	AC	GT	T	GA
	Section 22								
(1597)	1597	1610	1620	1630	1640	1650	1660	1670	1672
GagMod_SF2(1510)	--								
GagProtMod_SF2(GP1)(1597)	AGCCCAAAGATGATCGGGGGATCATCAAGGTGGGGCTTCATCAAGGAGCAGATCCCCTGGAGATCTGGGG								
GagProtMod_SF2(GP2)(1597)	AACCAAAAATGATAGGGGGATCGGGGGCTTCATCAAGGTGGGCACTACGACAGATACTGTAGAAATCTGTGG								
Consensus(1597)	A	CC	AA	ATGAT	GG	GGGATCGGGGCTTCATCAAGGTGG	GGCAGTACGACAGAT	CC	GT
	GA	ATCTG	GG	AT	AT	GC	AC	CC	GT
	Section 23								
(1673)	1673	1680	1690	1700	1710	1720	1730	1740	1748
GagMod_SF2(1510)	--								
GagProtMod_SF2(GP1)(1673)	CCACAAAGGCATCGGCCACCGTGTGGGGCCCCACCCCGTGAACATCATCGGCCAACCTGCTGACCCAGATC								
GagProtMod_SF2(GP2)(1673)	ACATAAAGCTATAGGTACAGTTAGTAGGACCTACACTGTCAACATAATTGGAAGAATCTGTGACCCAGATC								
Consensus(1673)	CA	AA	GC	AT	GG	AC	GT	T	GT
	GG	CC	AC	CC	AC	GT	AAACAT	AT	GG
	Section 24								
(1749)	1749	1760	1770	1780	1790	1800	1810	1820	1824
GagMod_SF2(1510)	--								
GagProtMod_SF2(GP1)(1749)	GGCTGCACCTGAACTCCCCATCGAGACGGTGGCCCGTGAAGCTGAAGCGGGGATGGACGGCCCA								
GagProtMod_SF2(GP2)(1749)	GGCTGCACCTGAACTCCCCATCGAGACGGTGGCCCGTGAAGCTGAAGCGGGGATGGACGGCCCA								
Consensus(1749)	GGCTGCACCTGAACTCCCCATCGAGACGGTGGCCCGTGAAGCTGAAGCGGGGATGGACGGCCCA	AT	GAGACGGTGGCCCGTGAAG	TGAAGCGGGGATGGACGGCCCA					
	Section 25								
(1825)	1825	1830	1847						
GagMod_SF2(1510)	--								
GagProtMod_SF2(GP1)(1825)	AGGTCAAGGAGTGGGCCCTGTAA								
GagProtMod_SF2(GP2)(1825)	AGGTCAAGGAAATGGCCATTGTAA								
Consensus(1825)	AGGTCAAGGAAATGGCCATTGTAA	TGGCC	TGTAA						

FIG. 80E

131 / 131

TataminoSF162.opt

ATGGAGCCGTTGGACCCCCGCCCTGGAAAGCACCCCGGCAGGCCAGCCCCAA
GACCGCCCTGGACCAAACTGGCTACTTGCAGAAGTGTGCTGCTTCCACAGGTGTGCTT
CATCACCAAGGGCCTGGGCATCAGCTACGGCCGGCAAGAAAGGCCAGGCCAGGCCAGGCC

FIG. 81
(SEQ ID NO:89)

Tat_Cys22_SF162

MEPVDPRLEPWKHPGSQPKTAGTNCYCKKCCFHQCQVCFITKGGLISYGRKKRQRRAAPPDSE
VHQVSLPKQPASQPQGDPTGPKEKKVERETETDPVHZ

FIG. 82
(SEQ ID NO:90)

SEQUENCE LISTING

<110> Chiron Corporation

<120> IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
OF VIRUS-LIKE PARTICLES

<130> 1621.100

<140>

<141>

<160> 90

<170> PatentIn Ver. 2.0

<210> 1

<211> 1509

<212> DNA

<213> Human immunodeficiency virus

<400> 1

atgggtgcga gagcgtcggt attaagcggg ggagaattag ataaatggga aaaaattcgg 60
ttaaggccag ggggaaagaa aaaatataag ttaaaacata tagtatggc aagcagggag 120
ctagaacgat tcgcagtcaa tcctggctg ttagaaacat cagaaggctg cagacaata 180
ttgggacagc tacagccatc cttcagaca ggatcagaag aacttagatc attatataat 240
acagttagcaa ccctctattg tgtacatcaa aggatagatg taaaagacac caaggaagct 300
ttagagaaga tagaggaaga gcaaaacaaa agtaagaaaa aggcacagca agcagcagct 360
gcagctggca caggaaacag cagccaggc agccaaaatt accctatagt gcagaaccta 420
cagggcaaa tggtacatca gcccataatca cctagaactt taaatgcattt ggtaaaagta 480
gtagaagaaa aggctttcag cccagaagta atacccatgt tttcagcatt atcagaagga 540
gccacccac aagatttaaa caccatgcta aacacagtgg ggggacatca agcagccatg 600
caaatgttaa aagagactat caatgaggaa gctcagaat gggatagagt gcatccagt 660
catgcaggc ctattgcacc aggc当地atg agagaaccaa ggggaagtga catagcagga 720
actactagta cccttcagga acaaataatgaa tggatgacaa ataatccacc tatccagta 780
ggagaaatct ataaaagatg gataatcctg ggattaaata aaatagtaag aatgtatagc 840
cctaccagca ttctggacat aagacaagga ccaaaggaac ccttttagaga ttatgttagac 900
cggttctata aaactctaag agccgaacaa gcttcacagg atgtaaaaaa ttggatgaca 960
gaaaccttgt tggccaaaa tgccaaaccca gattgtaaat ctatTTTaaa agcattggga 1020
ccagcagcta cactagaaga aatgtatgaca gcatgtcagg gagtggggg accccggccat 1080
aaagcaagag ttttggctga agccatgagc caagtaacaa atccagctaa cataatgatg 1140
cagagaggca attttaggaa ccaaagaaaag actgttaagt gtttcaattt tggcaagaa 1200
gggcacatag cccaaaattt cagggccctt aggaaaaagg gctgttgag atgtgaaagg 1260
gaaggacacc aatgtaaatgaa ttgcactgag agacaggcta attttttagg gaagatctgg 1320
ccttcctaca agggaaaggcc agggaaattt cttagagca gaccagagcc aacagcccc 1380
ccagaagaga gcttcaggtt tggggaggag aaaacaactc cctctcagaa gcaggagccg 1440
atagacaagg aactgtatcc tttaacttcc ctcagatcac tctttggcaa cgaccctcg 1500
tcacaataa 1509

<210> 2

<211> 1845

<212> DNA

<213> Human immunodeficiency virus

<400> 2

atgggtgcga gagcgtcggt attaagcggg ggagaattag ataaatggga aaaaattcgg 60
ttaaggccag ggggaaagaa aaaatataag ttaaaacata tagtatggc aagcagggag 120
ctagaacgat tcgcagtcaa tcctggctg ttagaaacat cagaaggctg cagacaata 180

ttgggacagc tacagccatc ccttcagaca ggatcagaag aacttagatc attatataat 240
 acagtagcaa ccctctattt tgtacatcaa aggatagatg taaaagacac caaggaagct 300
 ttagagaaga tagaggaaga gcaaaacaaa agtaagaaaa aggcacagca agcagcagct 360
 gcagctggca caggaaacag cagccaggc accaaaatt accctatagt gcagaaccta 420
 cagggcataa tggtacatca ggcataatca cctagaactt taaatgcatt ggtaaaagta 480
 gtagaagaaa aggcttcag cccagaagta ataccatgt tttcagcatt atcagaagga 540
 gccacccac aagattnaa caccatgcta aacacagtgg ggggacatca agcagccatg 600
 caaatgttaa aagagactat caatgaggaa gctgcagaat gggatagatg gcatccagtg 660
 catgcaggc ctattgcacc aggc当地 aagagaaccaa ggggaagtga catagcagga 720
 actactagta cccttcagga acaaataatg tggatgacaa ataatccacc tatcccagta 780
 ggagaaatct ataaaagatg gataatcctg ggattaaata aaatagtaag aatgtatagc 840
 cctaccagca ttctggacat aagacaagga ccaaaggAAC cctttagaga ttatgttagac 900
 cggttctata aactctaag agccgaaccaa gcttcacagg atgtaaaaaa ttggatgaca 960
 gaaaccttgt tggtccaaaa tgcaaaacca gattgtaaatg ctatTTTAA agcattggga 1020
 ccagcagcta cactagaaga aatgtatgaca gcatgtcagg gatgtggggg accccggccat 1080
 aaagcaagag ttttggctga agccatgagc caagtaacaa atccagctaa cataatgtatg 1140
 cagagaggca attttagaa ccaaagaaaactg tttcaattt gggcaaaagaa 1200
 gggcacatag cccaaaattt cagggccctt agggaaaagg gctgttggag atgttggagg 1260
 gaaggacacc aatgttggggg ttgcactgag agacaggcta attttttagg gaagatctgg 1320
 ctttcctaca agggaaattttt cttcagagca gaccagagcc aacagccccca 1380
 ccagaagaga gcttcaggtt tggggaggag aaaacaactc ccttcagaa gcaggagccg 1440
 atagacaagg aactgtatcc tttaacttcc ctcagatcac tctttggcaatcg 1500
 tcacaataag gatagggggg caactaaagg aagcttattt agatacagga gcagatgata 1560
 cagtattttaga agaaatgtatg ttgcaggaa aatggaaacc aaaaatgtatg gggggattt 1620
 gaggttttat caaagttaaga ctagtacatc agataccatgt agaaatctgt ggacataaag 1680
 ctataggtac agtatttagta ggacctacac ctgtcaacat aatttggaa aatctgttga 1740
 ctcagattgg ttgtacttta aatttccccca ttagtcttat taaaactgtt ccagtaaaat 1800
 taaagccagg aatggatggc ccaaagttt agcaatggcc attgtatg 1845

<210> 3

<211> 4313

<212> DNA

<213> Human immunodeficiency virus

<400> 3

atgggtgcga gagcgtcggt attaagcggg ggagaattttag ataaatggga aaaaattcgg 60
 ttaaggccag ggggaaagaa aaaatataag taaaacata tagtatgggc aagcaggag 120
 ctagaacatc tcgcgtcaatc tcctggctg ttagaaacat cagaaggctg cagacaaata 180
 ttggacagc tacagccatc cttcagaca ggatcagaag aacttagatc attatataat 240
 acagtagcaa ccctctattt tgtacatcaa aggatagatg taaaagacac caaggaagct 300
 ttagagaaga tagaggaaga gcaaaacaaa agtaagaaaa aggcacagca agcagcagct 360
 gcagctggca caggaaacag cagccaggc accaaaattt accctatagt gcagaaccta 420
 cagggcataa tggtacatca ggcataatca cctagaactt taaatgcatt ggtaaaagta 480
 gtagaagaaa aggcttcag cccagaagta ataccatgt tttcagcatt atcagaagga 540
 gccacccac aagattnaa caccatgcta aacacagtgg ggggacatca agcagccatg 600
 caaatgttaa aagagactat caatgaggaa gctgcagaat gggatagatg gcatccagtg 660
 catgcaggc ctattgcacc aggc当地 aagagaaccaa ggggaagtga catagcagga 720
 actactagta cccttcagga acaaataatg tggatgacaa ataatccacc tatcccagta 780
 ggagaaatct ataaaagatg gataatcctg ggattaaata aaatagtaag aatgtatagc 840
 cttaccagca ttctggacat aagacaagga ccaaaggAAC cctttagaga ttatgttagac 900
 cggttctata aactctaag agccgaaccaa gcttcacagg atgtaaaaaa ttggatgaca 960
 gaaaccttgt tggtccaaaa tgcaaaacca gattgtaaatg ctatTTTAA agcattggga 1020
 ccagcagcta cactagaaga aatgtatgaca gcatgtcagg gatgtggggg accccggccat 1080
 aaagcaagag ttttggctga agccatgagc caagtaacaa atccagctaa cataatgtatg 1140
 cagagaggca attttagaa ccaaagaaaactg tttcaattt gggcaaaagaa 1200
 gggcacatag cccaaaattt cagggccctt agggaaaagg gctgttggag atgttggagg 1260
 gaaggacacc aatgttggggg ttgcactgag agacaggcta attttttagg gaagatctgg 1320
 ctttcctaca agggaaattttt cttcagagca gaccagagcc aacagccccca 1380
 ccagaagaga gcttcaggtt tggggaggag aaaacaactc ccttcagaa gcaggagccg 1440

<210> 4
<211> 1515

<212> DNA

<213> Art:

Protein sequences

223

<223> Description of Artificial Sequence: synthetic
HIV-Gag

<400> 4

gccaccatgg gcgcccgccc cagcgtgctg agcggcggcg agctggacaa gtgggagaag 60
 atccgcctgc gccccggcgg caagaagaag tacaagctga agcacatcggt gtgggccagc 120
 cgcgagctgg agcgcttcgc cgtgaacccc ggctgctgg agaccagcga gggctgcccgc 180
 cagatcctgg gccagctgca gcccagctg cagacccggca gcgaggagct ggcagcctg 240
 tacaacaccg tggccaccct gtactgcgtg caccagcgc tcgacgtcaa ggacaccaag 300
 gagggcctgg agaagatcga ggaggagcag aacaagtcca agaagaaggc ccagcaggcc 360
 gccgcgcgcg cccgcaccgg caacagcgc caggtgagcc agaactaccc catcggtcag 420
 aacctgcagg gccagatggt gcaccaggcc atcagcccc gcaccctgaa cgccctgggtg 480
 aagggtgtgg aggagaaggc cttcagcccc gaggtgatcc ccatgtttag cgccctgagc 540
 gagggcgcca ccccccaggaa cctgaacacg atgttgaaca cctggggcg ccaccaggcc 600
 gccatgcaga tgctgaagga gaccatcaac gaggaggccg cegagtggga ccgcgtgcac 660
 cccgtgcacg cccgcgcgc cagatgcgcg agccccgggg cagcagacatc 720
 gccggccacca ccagcaccct gcaggaggcag atcggctggta tgaccaacaa ccccccattc 780
 cccgtggcg agatctacaa gcggtgatc atcctggggc tgaacaagat cgtgcggatg 840
 tacagccccca ccagcatcct ggacatccgc cagggccccca aggagccctt ccgcgactac 900
 gtggaccgcgt tctacaagac cctgcgcgt gaggaggccca gcaggacgt gaagaactgg 960
 atgaccgaga ccctgcgtgt gcagaacgc aaccccgact gcaagaccat cctgaaggct 1020
 ctcggccccc cggccaccct ggaggagatg atgaccgcgc cccaggcggt gggcgcccc 1080
 ggccacaagg cccgcgtgt ggccgaggcg atgagccagg tgacgaaccc ggcgaccatc 1140
 atgatgcagg gggcaacctt ccgcaccccg cggaaagaccg tcaagtgcctt caactgcggc 1200
 aaggagggcc acaccggccag gaactgcgcg gccccccgcg agaaggcggt ctggcgctgc 1260
 ggccgcgagg gcaccaggat gaaggactgc accgagcgcg aggccaactt cctggcaag 1320
 atctggccca getacaaggg cccgcgcgcg aaccttcgtc agagccggcc cgagccacc 1380
 gcccccccg aggagagctt ccgcgttgcg gaggagaaga ccaccccccag ccagaagcag 1440
 gagcccatcg acaaggagct gtacccctg accagcctgc gcagcctgtt cgcaacgcac 1500
 cccagcagcc agtaa 1515

<210> 5

<211> 1853

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
HIV-Gag-protease

<400> 5

gccaccatgg gcgcccgccc cagcgtgctg agcggcggcg agctggacaa gtgggagaag 60
 atccgcctgc gccccggcgg caagaagaag tacaagctga agcacatcggt gtgggccagc 120
 cgcgagctgg agcgcttcgc cgtgaacccc ggctgctgg agaccagcga gggctgcccgc 180
 cagatcctgg gccagctgca gcccagctg cagacccggca gcgaggagct ggcagcctg 240
 tacaacaccg tggccaccct gtactgcgtg caccagcgc tcgacgtcaa ggacaccaag 300
 gagggcctgg agaagatcga ggaggagcag aacaagtcca agaagaaggc ccagcaggcc 360
 gccgcgcgcg cccgcaccgg caacagcgc caggtgagcc agaactaccc catcggtcag 420
 aacctgcagg gccagatggt gcaccaggcc atcagcccc gcaccctgaa cgccctgggtg 480
 aagggtgtgg aggagaaggc cttcagcccc gaggtgatcc ccatgtttag cgccctgagc 540
 gagggcgcca ccccccaggaa cctgaacacg atgttgaaca cctggggcg ccaccaggcc 600
 gccatgcaga tgctgaagga gaccatcaac gaggaggccg cegagtggga ccgcgtgcac 660
 cccgtgcacg cccgcgcgc cagatgcgcg agccccgggg cagcagacatc 720
 gccggccacca ccagcaccct gcaggaggcag atcggctggta tgaccaacaa ccccccattc 780
 cccgtggcg agatctacaa gcggtgatc atcctggggc tgaacaagat cgtgcggatg 840
 tacagccccca ccagcatcct ggacatccgc cagggccccca aggagccctt ccgcgactac 900
 gtggaccgcgt tctacaagac cctgcgcgt gaggaggccca gcaggacgt gaagaactgg 960
 atgaccgaga ccctgcgtgt gcagaacgc aaccccgact gcaagaccat cctgaaggct 1020
 ctcggccccc cggccaccct ggaggagatg atgaccgcgc cccaggcggt gggcgcccc 1080
 ggccacaagg cccgcgtgt ggccgaggcg atgagccagg tgacgaaccc ggcgaccatc 1140
 atgatgcagg gggcaacctt ccgcaccccg cggaaagaccg tcaagtgcctt caactgcggc 1200
 aaggagggcc acaccggccag gaactgcgcg gccccccgcg agaaggcggt ctggcgctgc 1260
 ggccgcgaaag gacaccaaattt gaaagattgc actgagagac aggctaattt tttaggaaag 1320

atctggcctt cctacaaggg aaggccaggg aattttcttc agagcagacc agagccaaca 1380
 gccccaccag aagagagctt caggttggg gaggagaaaa caactccctc tcagaagcag 1440
 gagccgatag acaaggaact gtatccctta acttccctca gatcactctt tggcaacgac 1500
 ccctcgac agtaaggatc ggcggccagc tcaaggaggc gctgctcgac accggcggcc 1560
 acgacaccgt gctggaggag atgaacctgc cggcaagtg gaagcccaag atgatcggcg 1620
 ggatcggggg cttcatcaag gtgcggcagt acgaccagat cccctggag atctcggcc 1680
 acaaggccat cggcacccgt ctggtgccca ccaccccccgt gaacatcatc ggcgcacaacc 1740
 tgctgaccca gatcggctgc accctgaact tccccatcag ccccatcgag acgggtggcc 1800
 tgaagctgaa gccggggatg gacggcccca aggtcaagca gtggccctg taa 1853

<210> 6
<211> 4319
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic
HIV-Gag-polymerase

<400> 6
gccaccatgg gcgcccgcc cagcgtgctg agcggccggcg agctggacaa gtgggagaag 60
atccgcctgc gccccggcg caagaagaag tacaagctga agcacatctt gtggccagc 120
cgcgagctgg agcgcttcgc cgtgaaccccc ggctgtctgg agaccagcga gggctgccc 180
cagatccctgg gccagctgca gcccagctg cagaccggca gcgaggagct gcgacgcctg 240
tacaacacccg tggccacccct gtactgctg caccagcgc tgcacgtcaa ggacaccaag 300
gaggccctgg agaagatcga ggaggagcag aacaagtcca agaagaaggc ccagcaggcc 360
gccgcgcggc cccgcaccgg caacagcgc caggtgagcc agaactaccc catctgcag 420
aacctgcagg gccagatggt gcaccaggcc atcagccccc gcacccctgaa cgcctgggtg 480
aagggtgtgg aggagaaggc cttcagccccc gaggtgatcc ccatgttcag cgccctgagc 540
gagggcgcca ccccccaggaa cctgaacacg atgttgaaca ccgtggccg ccaccaggcc 600
gccatgcaga tgctgaagga gaccatcaac gaggaggccg ccgagtggga ccgcgtgcac 660
cccggtgcacg cccgcggccat cggccggcc cagatgcgcg agcccccggc cagcgacatc 720
gccggcaccac ccagcacccct gcaggaggcag atcggctgga tgaccaacaa ccccccatac 780
cccggtggcg agatctacaa gcgggtggatc atcctggcc tgaacaagat cgtcgccatg 840
tacagcccca ccagcatccct ggacatccgc cagggcccca aggagccctt ccgcgactac 900
gtggaccgct tctacaagac cctgcgcgt gaggcaggcc gccaggacgt gaagaactgg 960
atgaccgaga ccctgctggt gcagaacgccc aaccccgact gcaagaccat cctgaaggct 1020
ctcgcccccg cggccacccct ggaggagatg atgaccgcct gccaggccgt gggccggcccc 1080
ggccacaagg cccgcgtgt ggccgaggcg atgagccagg tgacgaaccc ggcgaccatc 1140
atgatgcagc gcggcaactt ccgcaccccg cggaaagaccg tcaagtgtt caactgcggc 1200
aaggaggggcc acaccgccc gaaactgccc gccccccgca agaagggtgt ctggcgctgc 1260
ggccgcgaag gacaccaaatt gaaagattgc actgagagac aggctaattt ttttagggaaag 1320
atctggcctt cctacaaggg aaggccaggg aattttcttc agagcagacc agagccaaca 1380
gccccaccag aagagagctt caggttggg gaggagaaaa caactccctc tcagaagcag 1440
gagccgatag acaaggaact gtatccctta acttccctca gatcactctt tggcaacgac 1500
ccctcgac agtaaggatc ggcggccagc tcaaggaggc gctgctcgac accggcggcc 1560
acgacaccgt gctggaggag atgaacctgc cggcaagtg gaagcccaag atgatcggcg 1620
ggatcggggg cttcatcaag gtgcggcagt acgaccagat cccctggag atctcggcc 1680
acaaggccat cggcacccgt ctggtgccca ccaccccccgt gaacatcatc ggcgcacaacc 1740
tgctgaccca gatcggctgc accctgaact tccccatcag ccccatcgag acgggtggcc 1800
tgaagctgaa gccggggatg gacggcccca aggtcaagca gtggccctg accgaggaga 1860
agatcaaggc cctgtgtggag atctgcaccc agatggagaa ggagggcaag atcagcaaga 1920
tcggccccca gaaacccctac aacaccccg tggatccat caagaagaag gacagcacca 1980
agtggcgca gctgggtggac ttccgcgc tgaacaagcg cacccaggac ttctgggagg 2040
tgcagctggg catcccccac cccgcggcc tgaagaagaa gaagagcgtg accgtgctgg 2100
acgtggcgca cgcctacttc agcgtggccc tggacaagga cttccgcac tacaccgcct 2160
tcaccatccc cagcatcaac aacgagaccc cggcatccg ctaccagtag aacgtgctgc 2220
cccaggctg gaagggcage cccgcacatct tccagagcag catgaccaag atccctggagc 2280
ccttcgcaca gacatcgta tctaccagta catggacgac ctgtacgtgg 2340

gcagcgacct ggagatcggc cagcacccgca ccaagatcga ggagctgcgc cagcacctgc 2400
 tgcgtctgggg cttcaccaccc cccgacaaga agcaccaggaa ggagcccccc ttccctgtgaa 2460
 tgggtacga gtcgcacccc gacaagtggc ccgtgcagcc catcatgtg cccgagaagg 2520
 acagctggac cgtgaacgac atccagaagc tggtgggcaa gctgaactgg gccagccaga 2580
 tctacgccgg catcaaggtg aagcagctgt gcaagctgtc gcgcggcacc aaggccctga 2640
 ccgaggtgat cccctgacc gaggaggccg agctggagct ggccgagaac cgcgagatcc 2700
 tgaaggagcc cgtgcacgag gtgtactacg accccagcaa ggacctgtg gccgagatcc 2760
 agaaggcaggg ccagggccag tggacattacc agatctacca ggagcccttc aagaacctga 2820
 agaccggcaa gtacgccccg atgcggcgccc cacacccaa cgacgtgaag cagctgaccg 2880
 aggccgtgca gaagggtgagc accgagagca tcgtgatctg gggcaagatc cccaaagtta 2940
 agctgcccattt ccagaaggag acctgggagg cctgggtgat ggagtaactgg caggccaccc 3000
 ggatccccga gtgggagttc gtgaacaccc ccccccgtt gaagctgtgg taccagctgg 3060
 agaaggagcc catctgtggc gcccggaccc tctacgtggc cggcggccccc aaccgcgaga 3120
 ccaagctggg caaggccggc tacgtgaccg accgcggccg ccagaagggtg gtgagcatcg 3180
 ccgacaccac caaccagaag accgagctgc agccatcca cctggccctg caggacagcg 3240
 gcctggaggt gaacatctgt accgcacagcc agtacgcctt gggcatcata caggccccagc 3300
 ccgacaagag cgagagcggc ctggtagcc agatcatega cgacgtatc aagaaggaga 3360
 aggtgtaccc ggcctgggtg cccggccaca agggcatcg ggcaacaggc cagggtggaca 3420
 agctggtgag cggccggcatc cgcaagggtc tggctctgaa cggcatcgac aaggccccagg 3480
 aggagcacga gaagtaccac agcaactggc ggcctatggc cagcgactt aacctgcccc 3540
 ccgtgggtggc caaggagatc gtggccagct ggcacaaggc ggacgtgaag ggcggaggcca 3600
 tgcacggcca ggtggactgc agcccccgc tctggcagct ggactgcacc cacctggagg 3660
 gcaagatcat cctggtgccc gtgcacgtgg ccagcggcta catcgagcc gaggtgatcc 3720
 ccggccagac cggccaggag accgcctact tctctgtgaa gctggccggc cgctggcccg 3780
 tgaagaccat ccacaccgac aacggcagca acttaccatc caccaccgt aaggccgcct 3840
 gctggggc cggcatcaag caggagttcg gcatccctta caaccccccag agccaggcg 3900
 tggtgagag catgaacaac gagctgaaga agatcatcg ccaggtgcgc gaccaggccg 3960
 agcacctgaa gaccgcgtg cagatggccg tggctatcca caacttcaag cgcaaggcg 4020
 gcatcgccgg ctacagcgc ggcagcgc tctggacat catcgccacc gacatccaga 4080
 ccaaggagct gcagaaggcag atcaccacca tccagaactt ccgcgtgtac taccgcgaca 4140
 acaaggaccc cctgtggaa gggcccgcca agctgtgtg gaaggccgag ggcggcgatg 4200
 tgatccagga caacagcgcac atcaaggtgg tgcccccggc caaggccaa atcatccgcg 4260
 actacggcaa gcagatggcc ggcgacgact gctggccag ccgcccaggc gaggactag 4319

<210> 7

<211> 2031

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
HIV-Gag/HCV-core fusion polypeptide

<400> 7

gcccacatgg gcccggcgc cagcgtgtc agcggccggc agctggacaa gtgggagaag 60
 atccgcctgc gccccggcg caagaagaag tacaagctga agcacatgt gtggggccagc 120
 cgcgagctgg agcgatccgc cgtgaacccc ggcctgtgg agaccagcga gggctgcgc 180
 cagatccctgg gccagctgca gcccagctg cagacccggc gcgaggagct ggcagcctg 240
 tacaacaccg tggccaccct gtactgcgtc caccagcgc tgcacgtcaa ggacaccaag 300
 gagggccctgg agaagatcga ggaggaggc aacaagtcca agaagaaggc ccagcaggcc 360
 gcccgcggcc cccgcaccgg caacagcgc caggtgagcc agaactaccc catcgatcg 420
 aacctgcagg gccagatgtt gcaccaggcc atcagccccc gcaccctgtaa cgcctgggtg 480
 aagggtggtgg aggagaaggc cttcaggcccc gaggtgatcc ccatgttgc ggcctgtggc 540
 gagggccca ccccccggc cctgaacacg atgttgaaca ccgtggccgg ccaccaggcc 600
 gccatgcaga tgctgaaggc gaccatcaac gaggaggccg ccgagtggaa ccgcgtgcac 660
 cccgtgcacg cccggccccc cggccggcc cagatgcgcg agcccccggc cagcgacatc 720
 gcccggccacca ccagcaccct gcaggaggc atcggctgaa tgaccaacaa ccccccaccc 780
 cccgtggccg agatctacaa gcgggtggatc atcctggcc tgaacaagat cgtgcggatg 840
 tacagcccca ccagcatccgc caggccccca aggagccctt ccgcgactac 900

gtggaccgct tctacaagac cctgcgcgt gaggcaggcca gccaggacgt gaagaactgg 960
 atgaccgaga ccctgctggt gcagaacgcc aaccccgact gcaagaccat cctgaaggct 1020
 ctcggccccg cggccacccct ggaggagatg atgaccgcct gccaggcgt gggcggcccc 1080
 ggcacacaagg cccgcgtget ggcgcggcg atgagccagg tgacgaaccc ggcgaccatc 1140
 atgatgcage gcccgaacct cgcgaaccag cggaaagaccg tcaagtgcct caactgcggc 1200
 aaggaggggcc acaccgcag gaactgcgc gccccccgca agaaggcgt ctggcgctgc 1260
 ggcgcgagg gccaccagat gaaggactgc accgagcgc aggccaaacct cctgggcaag 1320
 atctggccca gctacaaggg cgcgcggc aacttccctgc agagccgccc cgagcccacc 1380
 gccccccccg aggagagctt cgcgttccggc gaggagaaga ccaccccgag ccagaagcag 1440
 gagcccatcg acaaggagct gtacccctgc accagcctgc gcagcctgtt cggcaacgcac 1500
 cccagcagcc agtcgacgaa tcctaaacact caaagaaaaa acaaacgtaa caccacccgt 1560
 cgcacacagg acgtcaagtt cccgggtggc ggtcagatcg ttggtggtt ttacttgggt 1620
 cgcgcgagg gcccctagatt gggtgtggc ggcgcgagaa agacttcga ggggtcgcaa 1680
 cctcgaggta gacgtcagcc tatccccaaag gctcgtcgcc cgcggcccgag gacctgggt 1740
 cagccgggtt acccttggcc cctctatggc aatgagggtt ggggtggc gggatggctc 1800
 ctgtctcccc gtggctctcg gcctagctgg ggccccacag accccccgcg taggtcgcc 1860
 aatttggta aggtcatcg tacccttacg tgccgttccg cgcacccat ggggtacata 1920
 ccgctcgctcg ggcgcctct tggaggcgct gccaggggccc tggcgcatgg cgtccgggtt 1980
 ctggaagacg gcgtgaacta tgcaacagg aacccttcgt gttgctcta g 2031

<210> 8

<211> 2025

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
HIV-Gag/HCV-Core fusion polypeptide

<400> 8

atgggtgcga gagcgtcggt attaagcggg ggagaatttag ataaatggg aaaaattcgg 60
 ttaaggccag ggggaaagaa aaaatataag taaaacata tagtatggc aagcaggag 120
 ctagaacat tcgcagtcaa tcctggctg ttagaaacat cagaaggctg cagacaaata 180
 ttgggacagc tacagccatc cttcagaca ggatcagaag aacttagatc attatataat 240
 acagtagcaa ccctctattt tgtacatcaa aggatagatg taaaagacac caaggaagct 300
 ttagagaaga tagaggaaga gcaaaacaaa agtaagaaaa aggcacagca agcagcagct 360
 gcagctggca caggaaacag cagccaggc agccaaaatt accctatagt gcagaaccta 420
 cagggcaaa tggtacatca ggcataatca cctagaacct taaatgcatt ggtaaaagta 480
 gtagaaagaaa aggctttcag cccagaagta atacccatgt tttcagcatt atcagaagga 540
 gcccacccac aagatttaaa caccatgta aacacagtgg ggggacatca agcagccatg 600
 caaatgttaa aagagactat caatgaggaa gctcagaat gggatagatg gcatccagt 660
 catgcaggc cttatgcacc aggccaaatg agagaacca gggaaagtga catagcagga 720
 actactatgt cccttcagga acaaataatgg tggatgacaa ataatccacc tatcccagt 780
 ggagaaatct ataaaagatg gataatccctg ggattaaata aaatagtaag aatgtatagc 840
 cttaccagca ttctggacat aagacaagga ccaaaggAAC ccttttagaga ttatgttagac 900
 cggttcttaa aactctaaag agccgaacaa gcttcacagg atgtaaaaaa ttggatgaca 960
 gaaacattgt tggtccaaaaa tgcaaaacca gattgtaaatg ctatTTTaaa agcattggg 1020
 ccagcagcta cactagaaga aatgatgaca gcatgtcagg gagtgggggg accccggccat 1080
 aaagcaagat ttttggctga agccatgagc caagtaacaa atccagctaa cataatgtatg 1140
 cagagaggca attttaggaa ccaaagaaag actgttaagt gtttcaatg tggcaaagaa 1200
 gggcacatag ccaaaaattt caggccccctt agaaaaaagg gctgttggag atgttggagg 1260
 gaaggacacc aaatgaaaga ttgcacttag agacaggctt attttttagg gaagatctgg 1320
 cttccctaca agggaaaggcc agggaaatttt cttcagagca gaccagagcc aacagcccc 1380
 ccagaagaga gcttcaggat tggggaggag aaaaacaactt cctctcagaa gcaggagccg 1440
 atagacaagg aactgtatcc tttaacttcc ctcagatcac tctttggcaaa cgaccctcg 1500
 tcacagtcga cgaatcctaa acctcaaaaga aaaaacaaac gtaacacccaa ccgtcgcccc 1560
 caggacgtca agttcccggtt tggcggtcag atcgttgggt gatgttactt gttgcccgc 1620
 agggggcccta gattgggtgt ggcgcgacg agaaagactt ccgagcggtc gcaacccctcg 1680
 ggtagacgtc agccctatccc caaggctcgat cggcccgagg gcaggacctg ggctcagccc 1740

gggtaccctt ggccctcta tggcaatgag ggctgcgggt gggcgggatg gtcctgtct 1800
 ccccgtggct ctcggcttag ctggggcccc acagaccccc ggcgttagtc gcgcatttg 1860
 ggttaaggta tcgataaccct tacgtgcggc ttgcggacc tcatgggta cataccgctc 1920
 gtcggcggcc ctcttgagg cgctggcagg gccctggcgc atggcgccg ggttctggaa 1980
 gacggcgtga actatgcaac agggaacctt cctggttgct cttag 2025

<210> 9
 <211> 1268
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: synthetic Gag
 common region

<400> 9
 gccaccatgg gcgcccgcgc cagcgtgctg agcggccgcg agctggacaa gtgggagaag 60
 atccgcctgc gccccggcgg caagaagaag tacaagctga agcacatgt gtggggccagc 120
 cgcgagctgg agcgcttcgc cgtaaacccc ggcctgctgg agaccagcga gggctgccgc 180
 cagatcctgg gccagctgca gcccagcctg cagacccggca gcgaggagct gcgcagcctg 240
 tacaacacccg tggccaccct gtactgcgtg caccagcgc tcgacgtcaa ggacaccaag 300
 gaggccctgg agaagatcga ggaggagcag aacaagtcca agaagaaggc ccagcaggcc 360
 gccgcgcgcg cccgcaccgg caacagcagc caggtgagcc agaactaccc catcgtcag 420
 aacctgcagg gccagatggt gcaccaggcc atcagccccc gcaccctgaa cgcctgggtg 480
 aagggtgtgg aggagaaggc cttcagccccc gaggtgatcc ccatgttcag cgccctgago 540
 gagggcgcga ccccccagga cctgaacacg atgttgaaca ccgtggcgg ccaccaggcc 600
 gccatgcaga tgctgaaggg gaccatcaac gaggaggccg ccgagtggga ccgcgtgcac 660
 cccgtgcacg cccgcgcgc cccgcgcgc cagatgcgcg agcccccgcg cagcgcacatc 720
 gccggcaccga ccagcaccct gcaggagcag atcggctgga tgaccaacaa ccccccattc 780
 cccgtggcgcg agatctacaa gcgggtggatc atcctgggccc tgaacaagat cgtgcggatg 840
 tacagcccca ccagcatccct ggacatccgc cagggccccc aggagccctt ccgcgactac 900
 gtggaccgct tctacaagac cctgcgcgc gaggcaggcc gccaggacgt gaagaactgg 960
 atgaccgaga ccctgcgtgt gcagaacgc aaccccgact gcaagaccat cctgaaggct 1020
 ctcggccccc cggccaccct ggaggagatg atgacccgcct gccaggcggc gggccgcgc 1080
 ggcacaagg cccgcgtgt ggccgaggccg atgagccagg tgacgaaccc ggcgaccatc 1140
 atgatgcagc gcccacaactt ccgcacccag cggaaagaccc tcaagtgtt caactgcggc 1200
 aaggagggcc acaccgcccag gaactgcgcg gccccccgca agaagggtg ctggcgctgc 1260
 ggcgcgca 1268

<210> 10
 <211> 20
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: HIV-Gag
 peptide p7G

<400> 10
 Gly Gly His Gln Ala Ala Met Gln Met Leu Lys Glu Thr Ile Asn Glu
 1 5 10 15

Glu Ala Ala Glu
 20

<210> 11
 <211> 30
 <212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer GAG5

<400> 11
aagaattcca tgggtgcgag agcgtcggt 30

<210> 12
<211> 30
<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer p55-SAL3

<400> 12
attcgtcgac tgtgacgagg ggtcggtgcc 30

<210> 13
<211> 34
<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer CORESAL5

<400> 13
atttgcgac gaatcctaaa cctcaaagaa aaac 34

<210> 14
<211> 30
<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer 173CORE

<400> 14
tattggatcc taagagcaac caggaaggtt c 31

<210> 15
<211> 21
<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer MS65

<400> 15
cgaccatcat ggatgcagcg c 21

<210> 16
<211> 30
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer MS66

<400> 16
aggattcgtc gagtcgctgc tggggtcgtt 30

<210> 17
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer XPANXNF

<400> 17
gcacgtggc ccggcgcc tc tagagc 26

<210> 18
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer XPANXNR

<400> 18
gctcttagagg cgccgggccc acgtgc 26

<210> 19
<211> 20
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: HIV p55 Gag
Major Homology Region

<400> 19
Asp Ile Arg Gln Gly Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg
1 5 10 15

Phe Tyr Lys Thr
20

<210> 20
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic p55
Gag Major Homology Region

<400> 20
gacatccgcc agggccccaa ggagcccttc cgcgactacg tggaccgctt ctacaagacc 60

<210> 21
<211> 15

<212> PRT

<213> Human immunodeficiency virus

<400> 21

Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg
1 5 10 15

<210> 22

<211> 5

<212> PRT

<213> Human immunodeficiency virus

<400> 22

Lys Ala Lys Arg Arg
1 5

<210> 23

<211> 4

<212> PRT

<213> Human immunodeficiency virus

<400> 23

Arg Glu Lys Arg
1

<210> 24

<211> 15

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: aa of
mut7.SF162 cleavage site

<400> 24

Ala Pro Thr Lys Ala Ile Ser Ser Val Val Gln Ser Glu Lys Ser
1 5 10 15

<210> 25

<211> 15

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: aa of
mut8.SF162 cleavage site

<400> 25

Ala Pro Thr Ile Ala Ile Ser Ser Val Val Gln Ser Glu Lys Ser
1 5 10 15

<210> 26

<211> 15

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: aa of
mut.SF162 cleavage site

<400> 26

Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Ser
1 5 10 15

<210> 27

<211> 15

<212> PRT

<213> Human immunodeficiency virus

<220>

<223> Description of Artificial Sequence: aa of native
cleavage site in US4

<400> 27

Ala Pro Thr Gln Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg
1 5 10 15

<210> 28

<211> 5

<212> PRT

<213> Human immunodeficiency virus

<220>

<223> Description of Artificial Sequence: aa of second
cleavage site in US4

<400> 28

Gln Ala Lys Arg Arg
1 5

<210> 29

<211> 15

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: aa of mut.US4
cleavage site

<400> 29

Ala Pro Thr Gln Ala Lys Arg Arg Val Val Gln Arg Glu Lys Ser
1 5 10 15

<210> 30

<211> 1419

<212> DNA

<213> Human immunodeficiency virus

<400> 30

gtagaaaaat tgggttgcac agtctattat ggggtacctg tggaaaga agcaaccacc 60
 actctatccc gtgcatacaga tgctaaagcc tatgacacag aggtacataa tgtctggcc 120
 acacatgcct gtgtacccac agaccctaac ccacaagaaa tagtattgaa aaatgtgaca 180
 gaaaattttt acatgtggaa aaataacatg ttagaacaga tgcatacgga tataatcagt 240
 ttatggatc aaagtctaa gcatgtgt aagttacccc cactctgt tactctacat 300
 tgcactaatt tgaagaatgc tactataacc aagagtagta attggaaaga gatggacaga 360
 ggagaaaataa aaaattgct tttcaaggtc accacaagca taagaaataa gatgcagaaa 420
 gaatatgcac tttttataa acttgatgt tgcatacgat ataatgataa tacaagctat 480
 aaattgtataa attgtacac ctcagtcatt acacaggct gtccaaaggt atcccttcaa 540
 ccaattccca tacattattt tgccccggct ggttttgcga ttctaaagtg taatgataag 600
 aagtcaatg gatcaggacc atgtacaaat gtcaacacag tacaatgtac acatggatt 660
 aggccagtag tgcactca attgtcttta aatggcagtc tagcagaaga agggtagta 720
 attagatctg aaaatttcac agacaatgtc aaaactataa tagtacagct gaaggatct 780
 gtagaaaaattt attgtacaaag acctaacaat aatacaagaa aaagtataac tataaggaccg 840
 gggagagcat tttatgcac aggagacata ataggagata taagacaacg acattgtac 900
 attatggag aaaaatggaa taacacttaa aaacagatag ttacaaaatt acaagcacaa 960
 ttggaaataa aaacaatagt cttaagcaa tcctcaggag gggaccacg aatttgcata 1020
 cacagttta attgtggagg ggaattttt tactgttaatt caacacagct ttttaatagt 1080
 acttggata atactatagg gccaataaac actaatggaa ctatcacact cccatgcaga 1140
 ataaaacaaa ttataaacag gtggcaggaa ttagggaaag caatgtatgc ccctccatc 1200
 agaggacaaa ttagatgctc atcaaatatt acaggactgc tattaacaag agatgggt 1260
 aaagagatca gtaacaccac cgagatctt acacctggag gtggagatg gagggacaat 1320
 tggagaagtg aattatataa atataaagta gtaaaaaattt agccatttgg agtagcaccc 1380
 accaaggcaa agagaagagt ggtgcacaga gaaaaaagag cagtgcacgct aggagctatg 1440
 ttccctgggt tcttggggac agcaggaaac actatggcg cacggtcact gacgctgacg 1500
 gtacaggccaa gacaattttt gtcgttata gtcacacagc agaacaattt gctgagagct 1560
 attgaggcgc aacagcatct gttgcaactc acagtctggg gcatcaagca gctccaggca 1620
 agagtcctgg ctgtggaaag atacctaag gatcaacacg tccttagggat ttgggttgc 1680

<210> 31

<211> 1932

<212> DNA

<213> Human immunodeficiency virus

<400> 31

gtagaaaaat tgggttgcac agtctattat ggggtacctg tggaaaga agcaaccacc 60
 actctatccc gtgcatacaga tgctaaagcc tatgacacag aggtacataa tgtctggcc 120
 acacatgcct gtgtacccac agaccctaac ccacaagaaa tagtattgaa aaatgtgaca 180
 gaaaattttt acatgtggaa aaataacatg ttagaacaga tgcatacgga tataatcagt 240
 ttatggatc aaagtctaa gcatgtgt aagttacccc cactctgt tactctacat 300
 tgcactaatt tgaagaatgc tactataacc aagagtagta attggaaaga gatggacaga 360
 ggagaaaataa aaaattgct tttcaaggtc accacaacgtaaataa gatgcagaaa 420
 gaatatgcac tttttataa acttgatgt tgcatacgat ataatgataa tacaagctat 480
 aaattgtataa attgtacac ctcagtcatt acacaggct gtccaaaggt atcccttcaa 540
 ccaattccca tacattattt tgccccggct ggttttgcga ttctaaagtg taatgataag 600
 aagtcaatg gatcaggacc atgtacaaat gtcaacacag tacaatgtac acatggatt 660
 aggccagtag tgcactca attgtcttta aatggcagtc tagcagaaga agggtagta 720
 attagatctg aaaatttcac agacaatgtc aaaactataa tagtacagct gaaggatct 780
 gtagaaaaattt attgtacaaag acctaacaat aatacaagaa aaagtataac tataaggaccg 840
 gggagagcat tttatgcac aggagacata ataggagata taagacaacg acattgtac 900
 attatggag aaaaatggaa taacacttaa aaacagatag ttacaaaatt acaagcacaa 960
 ttggaaataa aaacaatagt cttaagcaa tcctcaggag gggaccacg aatttgcata 1020
 cacagttta attgtggagg ggaattttt tactgttaatt caacacagct ttttaatagt 1080
 acttggata atactatagg gccaataaac actaatggaa ctatcacact cccatgcaga 1140
 ataaaacaaa ttataaacag gtggcaggaa ttagggaaag caatgtatgc ccctccatc 1200
 agaggacaaa ttagatgctc atcaaatatt acaggactgc tattaacaag agatgggt 1260
 aaagagatca gtaacaccac cgagatctt acacctggag gtggagatg gagggacaat 1320
 tggagaagtg aattatataa atataaagta gtaaaaaattt agccatttgg agtagcaccc 1380
 accaaggcaa agagaagagt ggtgcacaga gaaaaaagag cagtgcacgct aggagctatg 1440
 ttccctgggt tcttggggac agcaggaaac actatggcg cacggtcact gacgctgacg 1500
 gtacaggccaa gacaattttt gtcgttata gtcacacagc agaacaattt gctgagagct 1560
 attgaggcgc aacagcatct gttgcaactc acagtctggg gcatcaagca gctccaggca 1620
 agagtcctgg ctgtggaaag atacctaag gatcaacacg tccttagggat ttgggttgc 1680

tctggaaaac tcatttgcac cactgctgtg ctttggaaatg ctatgtggag taataaatct 1740
ctggatcaga tttggaataa catgacctgg atggagtggg agagagaaaa tgacaattac 1800
acaaaactaa tatacacctt aattgaagaa tcgcagaacc aacaagaaaa gaatgaacaa 1860
gaattttag aattggataa gtgggcagaat ttgttggaaatt ggtttgacat atcaaatacg 1920
ctgtggtata ta 1932

<210> 32
<211> 2457
<212> DNA
<213> Human

<400> 32

gtgaaaaaaat tggggtcac agtctattat ggggtacctg tgtggaaaga agcaaccacc 60
actctatttt gtgcatacga tgctaaagcc tatgacacag aggtacataa tgtctggcc 120
acacatgcct gtgtacccac agaccctaac ccacaagaaa tagtattgga aaatgtgaca 180
gaaaatttta acatgtggaa aaataaacatg gtagaacaga tgcatgagga tataatcagt 240
ttatggatc aaagtctaaa gccatgtgt aagttAACCC cactctgtgt tactctacat 300
tgcactaatt tgaagaatgc tactaatacc aagagttagta attggaaaga gatggacata 360
ggagaaaataa aaaattgctc tttcaaggc accacaagca taagaaataa gatgcagaaa 420
aatatgcac tttttataa acttgatgt gtagccaatag ataatgataa tacaagct 480
aaattgataa attgtAACAC ctcaagtcat acacaggcct gtccaaagggt atcctttgaa 540
ccaattccca tacattattg tgccccggct ggTTTGCAG ttctaaaggta taatgataag 600
aagttcaatg gatcaggacc atgtacaaat gtcagcacag tacaatgtac acatggaaatt 660
aggccagtag tgtcaactca attgtgttta aatggcagtc tagcagaaga aggggttaga 720
attagatctg aaaatttcac agacaatgcg aaaactataa tagtacagct gaaggatct 780
gtagaaaatta attgtacaag acctaacaat aatacaagaa aaagtataac tataggacccg 840
gggagagcat ttatgcaac aggagacata ataggagata taagacaagc acattgtaac 900
attagtggag aaaaatggaa taacacttta aaacagatag ttacaaaatt acaagcaca 960
tttgggataa aaacaatagt cttaagcaa tcctcaggag gggacccaga aattgtaatg 1020
cacagtttta attgtggagg ggaatttttc tactgttaatt caacacagct tttaatagt 1080
acttggaaaata atactatagg gccaaataac actaatggaa ctatcacact cccatgcaga 1140
ataaaacaaa ttataaacag gtggcaggaa gtagggaaag caatgtatgc ccctcccatc 1200
agaggacaaa ttagatgctc atcaaataatt acaggactgc tattaacaag agatggtgt 1260
aaagagatca gtaacaccac cgagatcttc agacctggag gtggagatata 1320
tggagaagtg aattatataa atataaaagta gtaaaaaatttgc agcattagg agtagcaccc 1380
accaaggccaa agagaagagt ggtgcagaga gaaaaaaagag cagtgcgcgtt 1440
ttccttgggt tcttgggagc agcagggagc actatggcg cacggtaact gacgctgacg 1500
gtacaggcca gacaattatt gtctggtata gtgcacagc agaacaattttgc 1560
attgaggccg aacagcatct gttcaactc acagtctggg gcatcaagca gctccaggca 1620
agagtccctgg ctgtggaaag atacctaaag gatcaacagc tccttagggat ttggggttgc 1680
tcgtggaaaac tcatttgcac cactgtgtg ccttggaaatg ctatggagtt 1740
ctggatcaga tttggataa catgacctgg atggagtggg agagagaaat taataaatttgc 1800
acaaaacttaa tatacacctt aattgaagaa tcgcagaacc aacaagaaaaa gaatgaacaa 1860
gaattatttag aattggataa gtgggcaagt ttgtggaaatt ggtttgacat atcaaaaatgg 1920
ctgtggata taaaaatatt cataatgata gtaggggtt tagtaggttt aaggatagg 1980
tttactgtgc tttctatagt gaatagagtt aggccaggat actcaccatt atcatttgc 2040
acccgcttcc cagccccaaag gggacccgac aggccccaaag gaatcgaaga agaagggtgga 2100
gagagagaca gagacagatc cagtccattt gtgcattgtt tattagcact catctggac 2160
atgtacgca gcctgtgcct cttcagctac caccgttga gagacttaat cttgatttgc 2220
gcgaggattt tggaacttctt gggacgcagg ggggtggaaag ccctcaagta ttggggaaat 2280
ttctgtcagt attggattca ggaactaaag aatagtgttgc ttatggat ttagttttgc 2340
gtatagcag tagctgaggg gacagatagg attatagaag tagcacaaag aattggtaga 2400
gttttctcc acataaccttag aagaataaga cagggtttt gaaaggctt qctataa 2457

<210> 33
<211> 1453
<212> DNA
<213> Arti

<220>

<223> Description of Artificial Sequence: gp120.modSF162

<400> 33

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccaag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccacccctg ttctgcgcca gcgacgccaa ggcctacgac 180
accgagggtgc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caaccccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtgtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
accccccctgt gcgtgaccct gcactgcacc aacctgaaga acgccccacca caccaagagc 420
agcaactgga aggagatgga cccgcccgcg atcaagaact gcagcttcaa ggtgaccacc 480
agcatccgca acaagatgca gaaggagtagc gcctgttct acaagctgga cgtgtgtccc 540
atcgacaacg acaacaccacg ctacaagctg atcaactgca acaccagcgt gatecaccacag 600
gcctgccccca aggtgagctt cgagccatc cccatccact actgcgcggcc cgccggcttc 660
gccatcctga agtgcaacga caagaagttc aacggcagcg gcccctgcac caacgtgagc 720
accgtgcagt gcacccacgg catccgcccc gtggtgagca cccagctgct gctgaacggc 780
agcctggcccg aggagggcgt ggtgatccgc agcgagaact tcaccgacaa cgccaaagacc 840
atcatcgtgc agctgaagga gagcgtggag atcaactgca cccgccccaa caacaacacc 900
cgcaagagca tcaccatcgg ccccgccgcgc gccttctacg ccaccggcga catcatcgcc 960
gacatccgccc aggcccactg caacatcgc ggcgagaagt ggaacaacac cctgaagcag 1020
atcgtgacca agctgcaggc ccagttcgcc aacaagacca tcgtgttcaa gcagagcagc 1080
ggcggcgcacc ccgagatcgat gatgcacagc ttcaactgct gggcgcgat ttctactgc 1140
aacagcaccc agctgttcaa cagcacctgg aacaacacca tcggccccaa caacaccaac 1200
ggcaccatca ccctgcctcg ccgcacatcaag cagatcatca accgctgca ggaggtggc 1260
aaggccatgt acgcccccccatcccgcc cagatccgcgt gcagcagacaa catcaccggc 1320
ctgtctgtga ccccgacggc cggcaaggag atcagcaaca ccaccgagat ttccggcccc 1380
ggcggcggcgc acatgcgcga caactggcgc agcgagctgt acaagtacaa ggttgtgaag 1440
atcgagccccc tgg 1453

```

<210> 34

<211> 1387

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp120.modSF162.delV2

<400> 34

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccaag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccacccctg ttctgcgcca gcgacgccaa ggcctacgac 180
accgagggtgc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caaccccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtgtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
accccccctgt gcgtgaccct gcactgcacc aacctgaaga acgccccacca caccaagagc 420
agcaactgga aggagatgga cccgcccgcg atcaagaact gcagcttcaa ggtgcccgc 480
ggcaagctga tcaactgcac caccagctgg atcacccagg cctgccccaa ggtgagcttc 540
gagcccatcc ccatccacta ctgcgcggcc gcccgttcg ccacatctgaa gtgcacacgc 600
aagaagttca acggcagcgg cccctgcacc aacgtgagca cctgtcagtg caccacggc 660
atccggcccg tggtgagcac ccagctgtgc ctgaacggc gcctggccga ggagggcgtg 720
gtgatccgca gcgagaactt caccgacac gccaagacca tcatcgtgca gctgaaggag 780
agcgtggaga tcaactgcac cccggcccaac aacaacaccc gcaagagacat caccatcgcc 840
cccgccgcgc ccttctacgc caccggcgc acatcggcgc acatccgcca ggcccaactgc 900
aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tctgtgacca gctgcaggcc 960
cagttcggca acaagaccat cgtgttcaag cagagcagcg gccggcgcaccc cgagatcgtg 1020
atgcacagct tcaactgcgg cggcgttgc ttctactgca acagcacca gctgttcaac 1080
agcacctgga acaacaccat cggccccaac aacaccaacg gcaccatcac cctgcctgc 1140

```

cgcatcaagc agatcatcaa ccgctggcag gaggtggca aggccatgt cgccccccc 1200
 atccgggcc agatccgtg cagcagcaac atcacggcc tgctgctgac ccgcgacggc 1260
 ggcaaggaga tcagcaacac caccgagatc ttccggcccg gccgcggca catgcgcgac 1320
 aactggcga gcgagctgt caagtacaag gtggtaaga tgcagccct gggcgtggcc 1380
 cccacca 1387

<210> 35
 <211> 1323

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 gp120.modSF162.delV1V2

<400> 35

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgctgt gctgtgtgg 60
 gcagtcctcg ttcggccag cgccgtggag aagctgtgg tgaccgtgt ctacggcg 120
 cccgtgtgga aggaggccac caccacccgt ttctgcgcca ggcacgcca ggcctacgac 180
 accgaggtgc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caacccccc 240
 gagatcgatgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtgtgg 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 accccctgt gcgtggcgc cggcaactgc cagaccagcg tgatcaccca ggcctgccc 420
 aaggtgagct tcgagccat ccccatccac tactgcgccc ccgcgggtt cgccatctg 480
 aagtgcacg acaagaagtt caacggcagc ggccctgtca ccaacgtgag caccgtgcag 540
 tgcacccacg gcatccgccc cgtggtgac acccagctgc tgctgaacgg cagctggcc 600
 gaggaggcgc tggtgatccg cagcgagaac ttcaaccgaca acgccaagac catcatcg 660
 cagctgaagg agagcggtgg aatcaactgc acccgccca acaacaacac ccgcaagagc 720
 atcaccatcg gccccggccg cgccttctac gccaccggcg acatcatcg cgcacatccgc 780
 caggcccact gcaacatcg cggcgagaag tggaaaacaaca ccctgaagca gatcggtgacc 840
 aagctgcagg cccagttcg cacaagacc atcgtttca agcagagcag cggcggcgac 900
 cccgagatcg tgatgcacag cttcaactgc ggccggcgagt tcttctactg caacagcacc 960
 cagctgttca acagcacccgt gaacaacacc atcggccca acaacaccaa cggcaccatc 1020
 accctgcct gccgcatcaa gcagatcatc aaccgtggc aggagggtggg caaggccatg 1080
 tacggccccc ccatccgcgg ccagatccgc tgcagcagca acatcacccg cctgtgtgt 1140
 acccgccacg gccggcaagga gatcagcaac accaccgaga tcttccgccc cggcggcgac 1200
 gacatgcgcg acaactggcg cagcgagctg tacaagtaca aggtggtgaa gatcgagccc 1260
 ctggcgtgg ccccccaccaa ggccaaagcgc cgcgtggtgc agcgcgagaa gcgctaactc 1320
 gag 1323

<210> 36

<211> 2025

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: gp140.modSF162

<400> 36

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgctgt gctgtgtgg 60
 gcagtcctcg ttcggccag cgccgtggag aagctgtgg tgaccgtgt ctacggcg 120
 cccgtgtgga aggaggccac caccacccgt ttctgcgcca ggcacgcca ggcctacgac 180
 accgaggtgc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caacccccc 240
 gagatcgatgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtgtgg 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 accccctgt gcgtgaccct gcaactgc aacctgaaga acgcccacca caccaagagc 420
 agcaactggaa aggagatgg aacggcgag atcaagaact gcagcttcaa ggtgaccacc 480
 agcatccgca acaagatgc aacggagtac gcctgttct acaagctgga cgtggtggccc 540
 atcgacaacg acaacacccag ctacaagctg acaaccagcgat gatcaccac 600

gcctccccca aggtgagctt cgagcccatc cccatccact actgcgcccc cgccggcttc 660
 gccatcctga agtgcacaga caagaagtgc aacggcagcg gcccctgcac caacgtgagc 720
 accgtgcagt gcacccacgg catccggccc gtggtgagca cccagctgt gctgaacggc 780
 agcctggccg aggagggcgt ggtatccgc agcgagaact tcaccgacaa cgccaagacc 840
 atcatcgtgc agctgaagga gagcgtggag atcaactgca cccgccccaa caacaacacc 900
 cgcaagagca tcaccatcgg ccccgccgc gccttctacg ccacceggcga catcatcgcc 960
 gacatccgcc aggccccactg caacatcagg ggcgagaagt ggaacaacac cctgaagcag 1020
 atcgtgacca agctgcaggc ccagttcggc aacaagacca tcgtgttcaa gcagagcagc 1080
 ggcggcgacc ccgagatcgt gatgcacagg ttcaactgcg gccggcgagtt ctctactgc 1140
 aacagcaccc agctgttcaa cagcacctgg aacaacacca tcggccccaa caacaccaac 1200
 ggcacccatca ccctgccccgt ccgcacatcaag cagatcatca accgctggca ggaggtggc 1260
 aaggccatgt acgcccccccc catccgcggc cagatccgct gcagcagcaa catcaccggc 1320
 ctgctgctga cccgcgacgg cggcaaggag attagcaaca ccacccgagat ctccggcccc 1380
 ggcggcgccg acatgcgcga caactggcgc agcgagctgt acaagtacaa ggtggtaag 1440
 atcgagcccc tgggcgtggc ccccaccaag gccaaggcgc gcgtggtgcg gcgcgagaag 1500
 cgcggcggtga ccctggcgcc catgttctgg gccttctgg gcgcggccgg cagcaccatg 1560
 ggcggcccgca gcctgaccct gaccgtgcag gcccggccagc tgctgagcgg catcgtcag 1620
 cagcagaaca acctgctgcg cgccatcgg gcccagcgc acctgctgca gctgaccgtg 1680
 tgggcatca agcagctgca ggcccgcgtg ctggccgtgg agcgctaccc gaaggaccag 1740
 cagctgctgg gcatctgggg ctgcagcggc aagctgatct gcaccaccc cgtgccccgg 1800
 aacgccagct ggagcaacaa gagcctggac cagatctggc acaacatgac ctggatggag 1860
 tgggagcgcg agatcgacaa ctacaccaac ctgatctaca ccctgatcga ggagagccag 1920
 aaccagcagg agaagaacgca gcaggagctg ctggagctgg acaagtggc cagcctgtgg 1980
 aactggttcg acatcagcaa gtggctgtgg tacatctaac tcgag 2025

<210> 37

<211> 1944

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.modSF162.delV2

<400> 37

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgctgt gctgtgtgga 60
 gcagtcttcg tttcgcccaag cgccgtggag aagctgtggg tgaccgtgtta ctacggcggtg 120
 cccgtgtgga aggaggccac caccacccctg ttctgcgcac gcgacgcacaa ggcctacgc 180
 accgaggtgc acaacgtgtt ggcacccac gcctgcgtgc ccacccgaccc caaccccccag 240
 gagatcgtgc tggagaacgt gaccgagaac ttcaacatgtt ggaagaacaa catggtgag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaaggccctg cgtaaagctg 360
 accccctgt gctgtgaccct gcaactgcaccc aacctgaaga acgcccaccaa caccaagagc 420
 agcaactgga aggagatggc cccgcggcgag atcaagaact gcagcttcaa ggtggcgcc 480
 ggcaagctga tcaactgcac caccagcgtg atcacccagg cctgccccaa ggtgagcttc 540
 gagcccatcc ccatccacta ctgcggcccc gccggcttcg ccacccctgaa gtgcaacgc 600
 aagaagttca acggcagcgg cccctgcaccc aacgtgagca ccgtgcagtg caccacccggc 660
 atccggcccg tggtgagcac ccagctgtg ctgaacggca gcctggccga ggagggcggtg 720
 gtgatccgca gcgagaactt caccgacaac gccaagacca tcatcgtgc gctgaaggag 780
 agcgtggaga tcaactgcac cccggccaccc aacaacaccc gcaagagcat caccatcgcc 840
 cccggcccg cccctctacgc caccggcgac atcatcgccg acatccgcac ggcccaactgc 900
 aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tctgtgacca gctgcaggcc 960
 cagttcggca acaagaccat cgtgttcaag cagagcagcg gccggcgaccc cgagatcgtg 1020
 atgcacagct tcaactgcgg cggcgatcc ttctactgc acagcaccac gctgttcaac 1080
 agcacttggaa acaacacccat cggcccaac aacaccaacg gcaccatcac cctggccctgc 1140
 cgcatcaagc agatcatcaa cccgcgtggcag gaggtggca aggccatgtt cgcggcccccc 1200
 atccggcccg agatccgctg cagcagcaac atcaccggcc tctgtgtgc gccgcacggc 1260
 ggcaaggaga tcagcaacac caccgagatc ttccggcccc gccggcgagca catgcgcgac 1320
 aactggcgca gcgagctgtt caagtacaag gtggtgaaaga tctggccccctt gggcggtggc 1380
 cccaccaagg ccaagcgccg cgtggctgcag cgcgagaagc gccgcgtgc gctggccgc 1440

atgttcctgg gtttcctggg cgccgccggc agcaccatgg gcgcggcag cctgaccctg 1500
 accgtgcagg cccggcagct gctgagcggc atcgtgcagc agcagaacaa cctgctgcgc 1560
 gccatcgagg cccagcagca cctgctgcag ctgaccgtgt ggggcatcaa gcagctgcag 1620
 gcccgcgtgc tggccgtggg ggcgtacactg aaggaccagc agctgctggg catctggggc 1680
 tgcagcggca agctgatctg caccacccgc gtgcctggg acgcgcagctg gagcaacaag 1740
 agcctggacc agatctggaa caacatgacc tggatggagt gggagcgcga gatcgacaac 1800
 tacaccaacc tgatctacac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
 caggagctgc tggagctggg caagtggggc agcctgtggg actggttcga catcagcaag 1920
 tggctgttgtt acatctaact cgag 1944

<210> 38

<211> 1944

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.modSF162.delV1/V2

<400> 38

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgctgct gctgtgtgg 60
 gcagtcctcg ttccggccag cggcgtggag aagctgtggg tgaccgtgtta ctacggcggt 120
 cccgtgtggg aggaggccac caccacccctg ttctgcgcac ggcgcaccaa ggcctacgac 180
 acccgagggtgc acaacgtgtt ggcgcacccac gcctgcgtgc ccaccgcaccc caacccccag 240
 gagatcggtgc tggagaacgtt gaccgagaac ttcaacatgtt ggaagaacaa catgtgtggag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
 acccccccgtgt gcgtgaccctt gcactgcacc aacctgaaga acgcacccaa caccaggagc 420
 agcaacttggg aggagatggg ccgcggcgag atcaagaactt gcagcttcaa ggtggcgcc 480
 ggcaagctgtca tcaactgcac caccacgtgtt atcacccagg cctgccccaa ggtgagcttc 540
 gagcccatcc ccatccacta ctgcgcggcc gccggcttcg ccacccctgtt gtcacacgac 600
 aagaagttca acggcagcgg cccctgcacc aacgtgagca ccgtgcagtg caccacccggc 660
 atccggcccg tggtgagcac ccagctgtgtt ctgaacggca gcctggccga ggagggcggt 720
 gtgatccgca gcgagaactt caccgacaac gccaagacca tcacgtgtca gctgaaggag 780
 agcgtggaga tcaactgcac ccgcggccaaac aacaacaccc gcaagagcat caccatcgcc 840
 cccggcccg ccttctacgc caccggcgac atcatcggtt acatccgcac ggccactgtc 900
 aacatcagcg gcgagaagtgg aacaacacc ctgaagcaga tcgtgacccaa gctgcaggcc 960
 cagttccggca acaagaccat cgtgttcaagt cagagcagcg gcggcgaccc cgagatcggt 1020
 atgcacagct tcaactgcgg cggcgagttt ttctactgtca acagcacccaa gctgttcaac 1080
 agcacttggg acaacaccat ccgcggccaaac aacaccaacg gcaccatcac cctgcccgtc 1140
 cgcacatcaagc agatcatcaa ccgtgcgtgtt cgggtggca aggccatgtt cgcggccggc 1200
 atccggcccg agatccgtgtt caccacggccat tgcgtgttgc cccgcacccggc 1260
 ggcaaggaga tcaactgcac caccgagatc ttccggcccg gcggcgccgacatgcgcgac 1320
 aactggcgca gcgagctgtt caagtacaag tgggtgaaga tggggccctt gggcggtggcc 1380
 cccaccaagg ccaagcgcggc cgtgggtgtt cggcgacccggc ggcgcgtgac cctggccggc 1440
 atgttcctgg gtttcctggg cgcgcggccagc agcaccatgg gggcccgaccc cctgaccctg 1500
 accgtgcagg cccggccagct gctgagcggc atcgtgcagc agcagaacaa cctgctgcgc 1560
 gccatcgagg cccagcagca cctgctgcag ctgaccgtgtt ggggcatcaa gcagctgcag 1620
 gcccgcgtgc tggccgtggg ggcgtacactg aaggaccagc agctgctggg catctggggc 1680
 tgcagcggca agctgatctg caccacccgc gtgcctggg acgcgcagctg gagcaacaag 1740
 agcctggacc agatctggaa caacatgacc tggatggagt gggagcgcga gatcgacaac 1800
 tacaccaacc tgatctacac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
 caggagctgc tggagctggg caagtggggc agcctgtggg actggttcga catcagcaag 1920
 tggctgttgtt acatctaact cgag 1944

<210> 39

<211> 2025

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.mut.modSF162

<400> 39

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccaag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccacccctg ttctgcgcca gcgacgccaa ggcctacgac 180
accgaggtgc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
accccccgt gctgtgaccct gcactgcacc aacctgaaga acgcccaccaa caccaagagc 420
agcaacttgg aaggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtgaccacc 480
agcatccgca acaagatgca gaaggagatgc gccctgttct acaagctgga cgtggtgccc 540
atcgacaacg acaacaccag ctacaagctg atcaactgca acaccagcgt gatcacccag 600
gcctggccca aggtgagctt cgagcccatc cccatccatc actgcgcccc cgccggcttc 660
gccatcttga agtgcacacg caagaagtgc aacggcagcg gccccctgcac caacgtgagc 720
accgtgcagt gcacccacgg catccggccc gtggtgagca cccagctgct gctgaacggc 780
agcctggcccg aggaggccgt ggtgatccgc akgagaact tcacccgacaa cgccaaagacc 840
atcatcgtgc agtgcacacg gagcgttgag atcaactgca cccggcccaa caacaacacc 900
cgcaagagca tcaccatcgcc cccggccgc gccttctacg ccacccggca catcatcgcc 960
gacatccgccc agggccactg caacatcagc ggcgagaagt ggaacaacac cctgaagcag 1020
atcggtgacca agtgcagggc ccagttcgcc aacaagacca tcgtgttcaa gcagagcagc 1080
ggcggcgacc cccggatcgat gatgcacacg ttcaactgca gggcgaggtt cttctactgc 1140
aacaggacccc agtgcgttca cagcacttgg aacaacacca tcggcccaa caacaccaac 1200
ggcaccatca ccctggccctg ccgcatacg acaatcatca accgctggca ggaggtggc 1260
aaggccatgt akgccccc catccggcgc cagatccgcgt gcagcagcaa catcaccggc 1320
ctgctgttga cccggcgcacgg cggcaaggag atcagcaaca ccacccgagat cttccggccc 1380
ggcggcgccg acatgcgcga caactggcgc akgagctgt acaagtacaa ggtggtgaag 1440
atcgagccccc tggggctggc cccaccaag gccaagcgc gctgtgtgca ggcgcgagaag 1500
agcgcgtga ccctggggcgc catgttctg ggttctgg ggcggccgg cagcaccatg 1560
ggcggccgcga gcctgaccct gaccgtgcag gcccggcagc tgctgagccg catcggtcag 1620
cagcagaaca acctgtgcgc cggccatcgag gcccggcagc acctgtgtca gctgacccgt 1680
tggggcatca agcagctgca gggccggcgtg ctggccgtgg agcgtactt gaaggaccag 1740
cagctgttgg gcatctgggg ctgcagcggc aagctgtatct gcaccacccgc cgtggccctgg 1800
aacggccagct ggagcaacaa gggcttgac cagatcttgg acaacatgac ctggatggag 1860
tggggcgccg agatcgacaa ctacaccaac ctgtatctaca ccctgtatcgaa ggagagccag 1920
aaccaggcagg agaagaacga gcaggagctg ctggagctgg acaagtggc cagcctgtgg 1980
aactggttcg acatcagcaaa gtggctgtgg tacatctaactcgag 2025

```

<210> 40

<211> 1944

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.mut.modSF162.delV2

<400> 40

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccaag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccacccctg ttctgcgcca gcgacgccaa ggcctacgac 180
accgaggtgc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
accccccgt gctgtgaccct gcactgcacc aacctgaaga acgcccaccaa caccaagagc 420
agcaacttgg aaggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtggcgcc 480
ggcaagctga tcaactgcaaa caccagcgtg atcaccacccagg cctggcccaa ggtgagcttc 540

```

gagcccatcc ccatccacta ctgcgcccc gcccggcttcg ccatcctgaa gtgcaacgcac 600
 aagaagttca acggcagcg cccctgcacc aacgtgagca ccgtgcagtg caccacgc 660
 atccgccccg tggtgagcac ccagctgtg ctgaacggca gcctggccga ggagggcgtg 720
 gtgatccgca gcgagaacctt caccgacaac gccaagacca tcatcgtgca gctgaaggag 780
 agcgtggaga tcaactgcac ccggcccaac aacaacaccc gcaagagcat caccatcgac 840
 cccggcccg cttctacgc caccggcgac atcatcgac acatccgca ggeccactgc 900
 aacatcagcg gcgagaagtg gaacaacacc ctgaaggcaga tcgtgaccaa gctgcaggcc 960
 cagttcggca acaagaccat cgtgttcaag cagagcagcg gcggcgaccc cgagatcg 1020
 atgcacagct tcaactgcgg cggcgagtgc ttctactgca acagcacca gctgttcaac 1080
 agcaccttgg acaacaccat cggcccaac aacaccaacg gcaccatcac cctgccc 1140
 cgcaccaaggc agatcatcaa ccgctggcag gaggtgggca aggcatgta cgccccccc 1200
 atccgcccccc agatccgctg cagcagcaac atcaccggcc tgctgctgac cccgcacggc 1260
 ggcaaggaga tcagcaacac caccgagatc ttccgccccg gcggcgccga catgcgegac 1320
 aactggcgca gcgagctgta caagtacaag gtggtaaga tcgagccctt gggcgtggcc 1380
 cccaccaagg ccaagcgccg cgtggcag cgcgagaaga gcgcgtgac cctggcgcc 1440
 atgttcttgg gtttcttgg cggcccgccg agacaccatgg gcgcggccgag cctgaccctg 1500
 accgtgcagg cccggccagct gctgagcgcc atcgctgcagc agcagaacaa cctgctgcgc 1560
 gccatcgagg cccagcagca cctgctgcag ctgaccgtgt ggggcataa gcagctgcag 1620
 gcccgcgtgc tggccgtggc ggcgtacccg aaggaccaggc agctgctggg catctgggc 1680
 tgcagcggca agctgatctg caccacccggttggc acgcacgtg gagcaacaag 1740
 agcctggacc agatctggaa caacatgacc tggatggagt gggagcgcga gatcgacaac 1800
 tacaccaacc ttagtctacac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
 caggagctgc tggagctggc caagtggcc accctgtgg actggttcga catcagcaag 1920
 tggctgtggt acatctaact cgag 1944

<210> 41

<211> 1836

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.mut.modSF162.delV1/V2

<400> 41

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtgtgt gctgtgtgg 60
 gcagtttcg tttcgccccag cgccgtggag aagctgtggg tgaccgtgta ctacggcg 120
 cccgtgtgg aaggaggccac caccacccctg ttctgcgcgc gcgcacccaa ggcctacgac 180
 accgaggtgc acaacgtgtg ggcacccac gcctgcgtgc ccaccgaccc caacccccag 240
 gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtgtgg 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 accccctgt gctggggcgc cggcaactgc cagaccaggc tgatcacccaa ggcctgcccc 420
 aagggtgagct tcgagcccat ccccatccac tactgcgcgc cgcgcggctt cgcacatcctg 480
 aagtgcacg acaagaagtt caacggcagc ggccccctgca ccaacgtgag caccgtgcag 540
 tgcacccacg gcatccgccc cgtggtgagc acccagctgc tgctgaacgg cagcctggcc 600
 gaggaggccg tggatccg cagcgagaac ttcaaccgaca acgccaagac catcatcg 660
 cagctgaagg agagcgtggc gatcaactgc acccgccccc acaacaacac cgcacagagc 720
 atcaccatcg gccccggccg cgccttctac gccaccggcg acatcatcg cgcacatccgc 780
 caggcccact gcaacatcag cggcgagaag tggaaacaaca ccctgaagca gatcgtgacc 840
 aagctgcagg cccagttccg caacaagacc atcgtttca acgcagagcag cggccggcgc 900
 cccgagatcg ttagtgcacag cttcaactgc ggccggcgtt tcttctactg caacagcacc 960
 cagctgttca acagcacccgt gaacaacacc atcggccccc acaacacccaa cggcaccatc 1020
 accctgcctt gccgcataa gcagatcatc aaccgctggc aggagggtggg caaggccatg 1080
 tacgcccccc ccatcccgccg ccagatccgc tgcagcagca acatcacccgg cctgctgctg 1140
 acccgccgacg gccgcagga gatcagcaac accaccggaa tcttccgccc cggccggccgc 1200
 gacatgcgcg acaactggcg cagcgagctg tacaagtaca aggtggtgaa gatcgagccc 1260
 ctggccgtgg ccccccacca ggcacggcgc cgcgtggc acgcgcgagaa gacgcggccgtg 1320
 accctggccg ccatgttccct gggcttccctg ggcgcggccg acgcacccat gggcgccccgc 1380
 accctgaccc tgaccgtgca ggcggcccg ctgctgagcg gcatcgtgca acgcacccatg 1440

aacctgctgc gcgccatcga ggcccagcag cacctgctgc agctgaccgt gtggggcatc 1500
 aagcaagtgc aggccccgcgt gctggccgtg gagcgctacc tgaaggacca gcagctgctg 1560
 ggcatctggg gctgcagcgg caagctgatc tgcaccaccc ccgtgccctg gaacgccagc 1620
 tggagcaaca agagecctgaa ccagatctgg aacaacatga cctggatgga gtgggagcgc 1680
 gagatcgaca actacaccaa cctgatctac accctgatcg aggagagcca gaaccagcag 1740
 gagaagaacg agcaggagct gctggagctg gacaagtggg ccagcctgtg gaactggttc 1800
 gacatcagca agtggctgtg gtacatctaa tcgag 1836

<210> 42

<211> 2025

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 gp140.mut7.modSF162

<400> 42

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgt gctgtgtgga 60
 gcagtcctcg ttccgcccag cgccgtggag aagctgtggg tgaccgtgtta ctacggcgtg 120
 cccgtgtgga aggaggccac caccacccctg ttctgcgcga ggcacgcggaa ggcctacgac 180
 accgaggtgc acaaactgtgtt ggccaccac ccctgcgtgc ccaccgcaccc caaccccccag 240
 gagatcgatc tggagaacatgtt gaccgagaac ttcaacatgtt ggaagaacaa catggtgtag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
 accccctgt gcgtgaccct gcactgcacc aacctgaaga acggccacca caccaagagc 420
 agcaactgga aggagatggg ccgcggcgag atcaagaact gcagcttcaa ggtgaccacc 480
 agcatccgca acaagatgca gaaggagtttgc ccctgttttca acaagctgga cgtggtgc 540
 atcgacaacg acaacaccacg ctacaagctg atcaactgtca acaccgcgt gatcaccacccag 600
 gcctggccca aggtgagctt cgagcccatc cccatccact actgcggcccg cggccggcttc 660
 gccatccgtga atgtcaacgca caagaagtgc aacggcagcg gcccctgcac caacgtgagc 720
 accgtgcagt gcacccacgg catccggccc gtgtgagca cccagctgt gctgaacggc 780
 agcctggccg aggagggcgt ggtgatccgc akgagaact tcaccgacaa cgccaaagacc 840
 atcatcgtgc agctgaagga gagegtggag atcaactgtca cccggcccaaa caacaacacc 900
 cgcaagagca tcaccatcgcc cccggccgc gccttctacg ccaccggcgca catcatcgcc 960
 gacatccgccc agggccactg caacatcgc ggcgagaagt ggaacaacac cctgaagcag 1020
 atcgtgatcca agctgcaggc ccagttcgcc aacaagacca tcgtgttcaa gcagagcgc 1080
 ggccggacc ccgagatcgt gatgcacagc ttcaactgtgc gccggcgagtt cttctactgc 1140
 aacagcaccc agctgttcaa cagcacctgg aacaacacca tcggcccaaa caacaccaac 1200
 ggcaccatca ccctggccctg cccgcataaag cagatcatca accgctggca ggaggtgggc 1260
 aaggccatgt acggccccc catccggccg cagatccgcgt gcagcagacaa catcacccggc 1320
 ctgctgctga cccgcgacgg cggcaaggag atcagcaaca ccaccggat cttccggccc 1380
 ggcggccggcg acatgcgcga caactggcgc agcgagctgt acaagtacaa ggtggtaag 1440
 atcgagccccc tggcgtggc ccccaccaag gccatcagca gcgtgggtgca gagcgagaag 1500
 agccgcgtga ccctggccgc catgttctg gggttctgg gcgcggccgg cagcaccatg 1560
 ggcggccgcga gcctgaccct gaccgtgcag gcccggcagc tgctgagcgg catcgtcag 1620
 cagcagaaca acctgtcgcg cgcgcacgc gcccggcagc acctgtcgcg gctgaccgtg 1680
 tggggcatca agcagctgca ggccggcggt ctggccgtgg agcgctacct gaaggaccag 1740
 cagctgctgg gcatctgggg ctgcagcggc aagctgatct gcaccaccgc cgtggccctgg 1800
 aacgcccacgt ggagcaacaa gagcctggac cagatctgg aacaacatgac ctggatggag 1860
 tgggagcgcg agatcgacaa ctacaccaaactgatctaca ccctgatcgaa ggagagccag 1920
 aaccagcagg agaagaacga gcaggagctg ctggagctgg acaagtggc cagcctgtgg 1980
 aactggttcg acatcagcaaa gtggctgtgg tacatctaactcgag 2025

<210> 43

<211> 1944

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.mut7.modSF162.delV2

<400> 43

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccaag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccacccctg ttctgcgcca gcgacgccaa ggcctacgac 180
accgaggtgc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caaccccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtgtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
accccccctgt gcgtgaccct gcactgcacc aacctgaaga acgcccccaa caccaagagc 420
agcaactgga aggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtgggccc 480
ggcaagctga tcaactgcaa caccagcgtg atcacccagg cctgccccaa ggtgagctte 540
gagcccatcc ccatccacta ctgcggggggcc gccggcttc ccatacctgaa gtgcaacgac 600
aagaagttca acggcagcgg cccctgcacc aacgtgagca ccgtgcagtg caccacggc 660
atccgccccg tggtagcac ccagctgctg ctgaacggca gcctggccga ggagggcgtg 720
gtgatccgca gcgagaactt caccgacaac gccaagacca tcatacgtaa gctgaaggag 780
agcgtggaga tcaactgcac ccgcggccaa aacaacaccc gcaagagcat caccatcgcc 840
cccgccgcgc ctttctacgc caccggcgac atcatcgccg acatccgca ggcggactgc 900
aacatcagcg gcgagaagtg gaacaacacc ctgaaggaga tcgtgaccaa gctgcaggcc 960
cagttcggca acaagaccat cgtgttcaag cagagcagcg gcggcgaccc cgagatcgtg 1020
atgcacagct tcaactgcgg cggcgagttc ttctactgca acagcacca gctgttcaac 1080
agcacctgga acaacaccat cggccggccaa aacaccaacg gcaccatcac cctgccccctg 1140
cgcatcaagc agatcatcaa ccgcgtggcgag gaggtgggca aggccatgta cgcccccccc 1200
atccgccccg agatccgctg cagcagcaac atcacccggcc tgctgctgac ccgcgcacggc 1260
ggcaaggaga ttagcaacac caccgagatc ttccggggcc gggggccgca catgcgcgac 1320
aactggcgca gcgagctgta caagtacaag gtggtagaaga tcgagccctt gggcgtaggc 1380
cccaccaagg ccatcagcg cgtggcgag aegagaaga ggcggcgtgac cctggggccc 1440
atgttcttgg gtttccctgggg cggccggccg agcaccatgg ggcggccgag cctgaccctg 1500
accgtgcagg cccggccagct gctgagcgcc atcgtgcagc agcagaacaa cctgctgcgc 1560
gccatcgagg cccagcagca cctgctgcag ctgaccgtgt ggggcataa gcagctgcag 1620
ggccggcgtgc tggccgtggg ggcgtacctg aaggaccagc agctgctggg catctggggc 1680
tgcagcggca agctgatctg caccacccggc gtggccctggg acgcccagctg gagcaacaag 1740
agcctggacc agatctggaa caacatgacc tggatggagt gggagcgcga gatcgacaac 1800
tacaccaacc tggatctacac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
caggagctgc tggagctgga caagtggcc accctgtggg actggttcga catcagcaag 1920
tggctgtggt acatctaact cgag 1944

```

<210> 44

<211> 1836

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.mut7.modSF162.delV1/V2

<400> 44

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccaag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccacccctg ttctgcgcca gcgacgccaa ggcctacgac 180
accgaggtgc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caaccccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtgtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
accccccctgt gcgtggggcgc cggcaactgc cagaccagcg tgatcacca ggcctggccc 420
aaggtgagct tcgagcccat ccccatccac tactgcggcc cggccggctt cgccatccctg 480
aagtgcacg acaagaagtt caacggcagc ggccctgtca ccaacgtgag caccgtgcag 540
tgcacccacg gcatccgccc cgtggtagc acccagctgc tgctgaacgg cagcctggcc 600
gaggagggcg tggtagcccg cagcgagaac ttccacccgaca acggccaaacatcatcgta 660

```

cagctgaagg agagcgtgga gatcaactgc accccggccca acaacaacac ccgcaagagc 720
 atcaccatcg gccccggccg cgccttctac gccacccggcg acatcatcg cgacatccgc 780
 caggcccact gcaacatcg cggcgagaag tggaaacaaca ccctgaagca gatcggtgacc 840
 aagctgcagg cccagttcg caacaagacc atcggttca agcagagcag cggccggcgc 900
 cccgagatcg tgatgcacag cttcaactgc ggccggcgagt tcttctactg caacagcacc 960
 cagctgttca acagcacctg gaacaacacc atccggccca acaacacccaa cggcaccatc 1020
 accctgcctt gccgcataa gcagatcata aaccgctggc aggagggtggg caaggccatg 1080
 tacggccccc ccatccgcgg ccagatccgc tgcaagcagca acatcaccgg cctgctgctg 1140
 acccggcggc gccggcaagga gatcagaaac accaccggaga tcttccggcc cggccggggc 1200
 gacatgcgcg acaactggcg cagcgagctg tacaagtaca aggtgggtgaa gatcgagccc 1260
 ctggcggtgg ccccccaccaa ggcacatcgc agcgtggtgc agagcgagaa gagcggccgtg 1320
 accctggcg ccatgttccct gggcttccctg ggcggccggc gcagcaccat gggcccccgc 1380
 agcctgaccc tgaccgtgca ggcggcccg ctgtgagcg gcategtgca gcagcagaac 1440
 aacctgtgc ggcgcatacg ggcggcccg cacatcgatc agctgaccgt gtggggcattc 1500
 aagcagctgc aggccccgcgt gctggccgtg gagcgttacc tgaaggacca gcagctgctg 1560
 ggcacatctggg gtcgcagcg caagctgtatc tgcacccaccg cctgtccctg gaacgccagc 1620
 tggagcaaca agagcctgga ccagatctgg aacaacatga cctggatgga gtgggagcgc 1680
 gagatcgaca actacacccaa cctgatctac accctgtatcg aggagagacca gaaccagcag 1740
 gagaagaacg agcaggagct gctggagctg gacaagtggg ccagcctgtg gaactgggttc 1800
 gacatcagca agtggctgtg gtacatctaa ctggag 1836

<210> 45

<211> 2025

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gpl40.mut8.modSF162

<400> 45

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
 gcagtcctcg ttcgccccag cgccgtggag aagctgtggg tgaccgtgta ctacggcggt 120
 cccgtgtgga aggaggccac caccacccctg ttctgcgcga ggcacgcacaa ggccttacgac 180
 acccgaggtgc acaacatcgatc ggcacccac gcctgcgtgc ccaccgcaccc caaccccccac 240
 gagatcgatc tggagaacatcg gaccgagaac ttcaacatgt ggaagaacaa catggtgag 300
 cagatcgacg aggacatcat cagcgtgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 acccccccgt gctgtaccctt gcactgcacc aacctgaaga acgcacccaa caccaagagc 420
 agcaactgga aggagatggc cgcggcgag atcaagaact gcagcttcaa ggtgaccacc 480
 agcatccgca acaagatgca gaaggagatc gcctgttct acaagctgga cgtgggtgccc 540
 atcgacaacg acaacaccacg ctacaagctg atcaactgca acaccagcgt gatcaccac 600
 gcctggccca agtgtgagctt cgagccatc cccatccact actgcgcctt cggccggcttc 660
 gccatcctga atgtcaacga caagaagttt aacggcagcg gcccctgcac caacgtgagc 720
 accgtgcactt gcaacccacgg catccggccc gtgtgagca cccagctgt gctgaacggc 780
 agcctggccg aggaggggcgt ggtgttccgc agcgagaact tcaccgacaa cgccaaagacc 840
 atcatcgatc agctgtggatc gacgtggatc atcaactgca cccggcccaaa caacaacacc 900
 cgcaagagca tcaccatcgg ccccgccgc gccttctacg ccaccggcga catcatccgc 960
 gacatccgca aggccccactg caacatcagc ggcgagaagt ggaacaacac cctgaagcag 1020
 atcgatcgatc agctgcggc ccagttccgc aacaagatcgatc tcgtgttcaa gcagagcagc 1080
 ggcggccacc cccagatcgatc gatcgacccgatc ttcactcgcc gccggcgagg tttctactgc 1140
 aacagcaccc agtggatcgatc cagcacctgg aacaacacca tggggcccaaa caacaccaac 1200
 ggcaccatca ccctggccctg cccgcatacg cagatcatcg accgtggca ggaggtgggc 1260
 aaggccatgt acggccccc catccggccg cagatccgt gcaagcagacaa catcaccggc 1320
 ctgtgtgtca cccgcgcacgg cggcaaggag atcagcaaca ccaccggat tttccggcc 1380
 ggcggccggc acatgcgcga caactggcgcc agcgagatcgatc acaagatcaca ggtgggtgaag 1440
 atcgatcgatc tggcgatcgatc cccaccatc gccatcgatc gcgtggatcgatc gagcggagaag 1500
 agcgcggcgatc ccctggccgc catgttccctg gggttccctg ggcggccggc cagcaccatg 1560
 ggcggccgc gctgtaccctt gaccgtgcag gcccggccgc tgctgagccgg catcgatcgatc 1620
 cagcagaaca acctgtgcgcg cggccatcgatc gcccggccgc acctgtgcgcg gctgaccgtg 1680

tggggcatca agcagctgca ggccgcgtg ctggccgtgg agcgctacct gaaggaccag 1740
 cagctgctgg gcatctgggg ctgcagcggc aagctgatct gcaccaccgc cgtgccctgg 1800
 aacgccagct ggagcaacaa gagcctggac cagatctgga acaacatgac ctggatggag 1860
 tgggagcgcg agatcgacaa ctacaccaac ctgatctaca ccctgatcga ggagagccag 1920
 aaccagcagg agaagaacga gcaggagctg ctggagctgg acaagtgggc cagctgtgg 1980
 aactggttcg acatcagcaa gtggctgtgg tacatcta ac tcgag 2025

<210> 46

<211> 1944

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp140.mut8.modSF162.delV2

<400> 46

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgctgt gctgtgtgg 60
 gcagtcctcg ttccgccccag cgccgtggag aagctgtgg tgaccgtgta ctacggcgtg 120
 cccgtgtggaa aggaggccac caccacccctg ttctgcgccttgcgcacaa ggcctacgac 180
 accgagggtgc acaacgtgtg ggcacccac gcctgcgtgc ccaccgaccc caaccccccag 240
 gagatcgtgc tggagaacgt gaccgagaac ttcaacatgtt ggaagaacaa catggtgag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 accccctgt gctgtgacccttgcactgcacc aacctgaaga acgccacccaa caccaggagc 420
 agcaacttggaa aggagatggaa ccgcggcgag atcaagaacttgcac gtagtttcaaa ggtggcgcc 480
 ggcaagcttgc tcaactgcac caccagcgtg atcaacccagg cctgcacccaa ggtgagcttc 540
 gagcccatcc ccatccacta ctgcgcaccc gccggcttcg ccatcctgaa gtgcaacgac 600
 aagaagttca acggcagcgg cccctgcacc aacgtgagca cctgtgcactg caccacggc 660
 atccggcccg tggtgagcac ccagctgtgc ctgaacggca gcttggccga ggaggcggtg 720
 gtgatccgca gogagaactt caccgacaac gccaagagcca tcatcgtgca gctgaaggag 780
 agcgtggaga tcaactgcac cccgcacccaa aacaacaccc gcaagagcat caccatcgcc 840
 cccggccgcg cttctacgc caccggcgac atcatcgccg acatccgcac gggccactgc 900
 aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tcgtgacccaa gctgcaggcc 960
 cagttcggca acaagaccat cgtgttcaag cagacgcgcg gccggcgcaccc cgagatcgtg 1020
 atgcacagct tcaactgcgg cggcgagttc ttctactgca acagcacccaa gctgttcaac 1080
 agcaccttggaa acaacaccat cggcccaac aacaccaacg gcaccatcac cctgcctgc 1140
 cgcataaagc agatcatcaa ccgtgtggcag gaggtggca aggccatgttgcgcaccc 1200
 atccggcccg agatccgctg cagcacaac atcaccggcc tgcgtgtgac cccgcacggc 1260
 ggcgaaggaga tcagcaacac caccgagatc ttccggcccg gccggcggcga catgcgcgc 1320
 aacttgcgcac gcgagctgttgc caagtacaag gtggtgaaga tcgagccctt gggcgtggcc 1380
 cccaccatcg ccatcagcgac cgtgtgtgcag agcgagaaga ggcgcgtgac cctggcgcc 1440
 atgttccctgg gtttccctggg cggccgcggc agcaccatgg ggcgcgcaccc cctgaccctg 1500
 accgtgcagg cccgcacccat gctgagcggc atcgtgcgcg acgagaacaa cctgctgcgc 1560
 gccatcgagg cccagcgcac cctgtgtgcag ctgaccgtgtt ggggcataaa gcagctgcag 1620
 gcccgcgtgc tggccgtggaa ggcgttacctg aaggaccgcg agctgtgtgg catctggggc 1680
 tgcagcggca agctgtatctg caccacccgcgttgcaccc gacgcgcaccc cctgaccctg 1740
 agccttggacc agatctggaa caacatgacc tggatggagt gggagcgcga gatcgacaac 1800
 tacaccaacc tcatctacac cctgtatcgac gagagccaga accagcaga gaagaacgag 1860
 caggagctgc tggagctggaa caagtggcc acgcctgtggaa actggttcga catcagcaag 1920
 tggctgtggatcatctaacttgcag 1944

<210> 47

<211> 1836

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp140.mut8.modSF162.delV1/V2

<400> 47

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
 gcagtcttcg tttcgcccaag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
 cccgtgtgga aggaggccac caccacccctg ttctgcgcca ggcacgccaa ggcctacgac 180
 accgaggtgc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caaccccccag 240
 gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 acccccctgt gcgtgggccc cggcaactgc cagaccagcg tgatcaccca ggcctgcccc 420
 aaggtgagct tcgagcccat ccccatccac tactgcgccc ccgcgggctt egccatcctg 480
 aagtgcacg acaagaagtt caacggcagg ggcctctgca ccaacgtgag caccgtgcag 540
 tgcacccacg gcatccgccc cgtggtgagc acccagctgc tgctgaacgg cagcctggcc 600
 gaggaggcgc tggtgatccg cagcgagaac ttacccgaca acgccaagac catcatcgtg 660
 cagctgaagg agagcgtgg aatcaactgc acccgccccca acaacaacac ccgcaagagc 720
 atcaccatcg gccccggccg cgccttctac gccaccggcg acatcatcg cgacatccgc 780
 caggcccact gcaacatcg cggcgagaag tggaaacaaca ccctgaagca gatcgtgacc 840
 aagctgcagg cccagttcg caacaagacc atcgtgttca agcagagcag cggcggcgc 900
 cccgagatcg tgatgcacag cttcaactgc ggcggcgagt tcttctactg caacagcacc 960
 cagctgttca acagcacctg gaacaacacc atcgccccca acaacaccaa cggcaccate 1020
 accctgcctt gccgcataa gcagatcatc aaccgctggc aggaggtggg caaggccatg 1080
 tacgcccccc ccatccgccc ccagatccgc tgacgcagca acatcaccgg cctgctgctg 1140
 acccgcgacg gccgcaggaa gatcagcaac accaccgaga tcttccggcc cggcggcggc 1200
 gacatgcgcg acaactggcg cagcgagctg tacaagtaca aggtggtaa gatcgagccc 1260
 ctggcgctgg cccccaccat cgccatcgc agcgtggtgc agagcgagaa gagcgcgtg 1320
 accctggcg ccatgttccct gggcttctgg ggcgcggccg gcagcaccat gggcggccgc 1380
 agcctgaccc tgaccgtgca ggcggccag ctgctgagcg gcatcgtgca gcagcagaac 1440
 aacctgtgc ggcgcatacg ggcggcagc cactctgtgc agctgaccgt gtggggcatc 1500
 aagcagctgc aggcccgcgt gctggcgctg gagcgctacc tgaaggacca gcagctgctg 1560
 ggcatctggg gctgcagcg ctagctgatc tgacaccaccc cctgcccggc gAACGCCAGC 1620
 tggagcaaca agagcctgg aacaacatga cctggatggg gtggagcgc 1680
 gagatcgaca actacaccaa cctgatctac accctgtatcg aggagagacca gaaccagcag 1740
 gagaagaacg agcaggagct gctggagctg gacaagtggg ccagcctgtg gaactggttc 1800
 gacatcagca agtggctgtg gtacatctaa ctcgag 1836

<210> 48

<211> 2547

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: gp160.modSF162

<400> 48

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
 gcagtcttcg tttcgcccaag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
 cccgtgtgga aggaggccac caccacccctg ttctgcgcca ggcacgccaa ggcctacgac 180
 accgaggtgc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caaccccccag 240
 gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 acccccctgt gcgtggccct gcaactgcacc aacctgaaga acgcccaccaa caccaagagc 420
 agcaactgg aaggatggg cggcgccag atcaagaact gcagcttcaa ggtgaccacc 480
 agcatccgca acaagatgcg gaaggagtagt gcccgttct acaagctgca cgtggcgccc 540
 atcgacaacg acaacacccag ctacaagctg atcaactgca acaccagcgt gatcaccac 600
 gcctggccca agtgagctt cgagccatc cccatccact actgcgcccc cggccggcttc 660
 gccatctgaa atgcacgca caagaagtgc aacggcagcg gccccctgcac caacgtgagc 720
 accgtgcagt gcaacccacgg catccggccc gtggtgagca cccagctgct gctgaacggc 780
 agccctggccg agagggcgt ggtgatccgc agcgagaact tcaccgacaa cgccaaagacc 840
 atcatcgtgc agtgcacgg gacgtggag atcaactgca cccggccca caacaacacc 900
 cgcaagagca tcaccatcg gccccggccg gccttctacg ccaccggcga catcatcggc 960
 gacatccgca agccactgca ggcgagaagt ggaacaacac cctgaagcag 1020

atcgtgacca agctgcaggc ccagttcggc aacaagacca tcgtgttcaa gcagagcagc 1080
 ggcggcgacc ccgagatcg tgcacacgc ttcaactcgc gccgcgagtt cttctactgc 1140
 aacagcaccc agctgttcaa cagcacctgg aacaacacca tcggcccca caacaccaac 1200
 ggcaccatca ccctgcctcg cccatcaag cagatcatca accgctgca ggaggtggc 1260
 aaggccatgt acggccccc catccgcggc cagatccgt gcagcagcaa catcaccggc 1320
 ctgctgtga cccgcgacgg cgcaaggag atcagcaaca ccaccgagat cttccgcccc 1380
 ggcggcgccg acatgcgcga caactggcgc agcgagctgt acaagtacaa ggtgtgaag 1440
 atcgagcccc tggcggtgg cccccaccaag gccaagcgc gcgtggtca ggcgagaag 1500
 cgcgcgtga ccctggcgc catgttctg ggcttcctgg ggcgcgcgg cagcaccatg 1560
 ggcggcccgca gcctgcaccc gaccgtgcag gcccgcgc tgctgagcgg catcgccag 1620
 cagcagaaca acctgtcg cgcgcgtgg gcccagcgc acctgtcga gctgaccgtg 1680
 tgggcattca acgagctgca ggcgcgtgg ctggccgtgg agcgctacct gaaggaccag 1740
 cagctgtgg gcatctgggg ctgcgcggc aagctgtatc gcaccacgc cgtggccctgg 1800
 aacgccagct ggagcaacaa gagcctggc cagatctgg acaacatgac ctggatggag 1860
 tgggagcgcg agatcgacaa ctacaccaac ctgatctaca ccctgatcga ggagagccag 1920
 aaccacgagg agaagaacga gcaggagctg ctggagctgg acaagtggc cagctgtgg 1980
 aactggttcg acatcagcaa gtggctgtgg tacatcaaga ttttcatcat gatcggtggc 2040
 ggcctggtgg gcctgcgcatt cgtgttcaacc gtgtgagca tcgtgaaccg cgtgcgcgcag 2100
 ggctacagcc ccctgagctt ccagacccgc ttcccccgc cccgcggccc cgaccgcggc 2160
 gagggcatcg aggaggaggg cggcgcgcgc gaccgcgc gcaagcggcc cctggtgac 2220
 ggcctgtgg ccctgatctg ggacgcctg cgcgcgtgt gcctgttccag ctaccacccg 2280
 ctgcgcgacc tgatcctgat cggccgcgc atcgtggagc tgctggccg cggccgtgg 2340
 gagggccctga agtactgggg caacctgtg cagtaactgg tccaggagct gaagaacagc 2400
 gccgtgagcc tgttcgacgc catgcgcattc gccgtggccg agggcaccga ccgcattatc 2460
 gaggtggccc agcgcatcg cggcgcctt ctgcacatcc cccgcgcatt cgcgcaggc 2520
 ttcgagcgcc ccctgctgtaa actcgag 2547

<210> 49

<211> 2466

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp160.modSF162.delV2

<400> 49

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgtgtgt 60
 gcagtttcg tttcgccca cgcgcgtggag aagctgtgg tgaccgtgtta ctacggcggt 120
 cccgtgtggaa aggaggccac caccacccctt ttctgcgcac ggcgcacccaa ggccttacgac 180
 accgaggtgc acaacgtgtt ggcgcacccac gcctgcgtgc ccaccgcaccc caaccccccag 240
 gagatcgatgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
 acccccccgt gctgtgaccct gcactgcacc aacctgaaga acgcacccaa caccaagagc 420
 agcaactggaa aggagatggc cgcgcggcgag atcaagaact gcaacttcaa ggtggccgc 480
 ggcaagctga tcaactgcac caccaggctg atcaccagg cctgcgcacca ggtgagctt 540
 gagcccatcc ccatccacta ctgcgcgcgc gccggcttc ccatcctgaa gtgcaacgc 600
 aagaagttca acggcagcgg cccctgcacc aacgtgagca cctgtcagtg caccacggc 660
 atccgcggcc tggtgagcac ccagctgtg ctgaacggca gcctggccga ggagggcggt 720
 gtgatccgca gcgagaactt caccgacaac gccaagacca tcatcggtca gctgaaggag 780
 agcgtggaga tcaactgcac cggcccccac aacaacaccc gcaagagcat caccatcgcc 840
 cccggccgcg ctttctacgc caccggcgac atcatcggtgc acatccgcac ggcgcactgc 900
 aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tcgtgaccaa gctgcaggcc 960
 cagttcgcca acaagaccat cgtgttcaag cagagcagcg gccgcgcaccc cgagatcggt 1020
 atgcacagct tcaactgcgg cggcgagttt ttctactgca acagcacccaa gctgttcaac 1080
 agcacttggaa acaacacccat cggcccccac aacaccaacg gcaccatcac cctggccctgc 1140
 cgcacatcaagc agatcatcaa cccgtggcag gaggtggcga aggccatgtaa cgcgcgcgc 1200
 atccggccgc agatccgtc cagcagcaac atcaccggcc tgctgtgac cgcgcacggc 1260
 ggcaaggaga tcagcaacac caccgagatc ttccggcccg gccgcggcga catcgccgac 1320

aactggcgca gcgagctgta caagtacaag gtggtgaaga tcgagccct gggcggtggc 1380
cccaccaagg ccaagcgccg cgtggtgcag cgcgagaagc gcgcgtgac cctggggcgcc 1440
atgttcctgg gtttcctggg cgccgcggc agcacccatgg gcgcgcgcag cctgaccctg 1500
accgtgcagg cccgcgcagct gctgagcgcc atcggtcagc agcagaacaa cctgctgcgc 1560
gccatcgagg cccagcagca cctgctgcag ctgaccgtgt ggggcataa gcagctgcag 1620
gcccgcgtgc tggccgtggc ggcgcatactg aaggaccagc agctgtggg catctggggc 1680
tgcagcgca agctgatctg caccacccgc gtgccttgga acgcgcgtg gagcaacaag 1740
agcctggacc agatctggaa caacatgacc tgatggagt gggagcgccg gatcgacaac 1800
tacaccaacc tgatctacac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
caggagctgc tggagctggc caagtggggc agcctgtggc actggttcga catcagcaag 1920
tggctgtggt acatcaagat cttcatctg atcggtggcg gcctgggtgg cctgcgcattc 1980
gtgttccaccc tgctgagcat cgtgaaccgc gtgcgcagg gctacagccc cctgagcttc 2040
cagacccgcg tccccggccc cgcggggggc gacgcggggc agggcatacgaa ggaggaggggc 2100
ggcgagcgccg acccgccaccc cagcagcccc ctggtgcacgc gcctgctggc cctgatctgg 2160
gacgacctgc gcagcctgtg cctgatcgac taccacccgc tgcgccgcct gatcctgtatc 2220
gcccggcgcgca tcgtggagct gctggggccgc cgccggctggg aggccctgaa gtactggggc 2280
aacctgtgc agtactggat ccaggagctg aagaacacgc ccgtgagct gttcgacgc 2340
atcgccatcg ccgtggccgcg gggcaccgcg cgcacatcg aggtggccca ggcacatcgcc 2400
cgccgccttc tgcacatccc cgcgcgcattc cgccagggtc tgcagcgccg cctgctgtaa 2460
ctcgag 2466

<210> 50

<211> 2358

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp160.modSF162.delV1/V2

<400> 50

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgg 60
gcagtcctcg ttcgcggccag cgccgtggag aagctgtggg tgaccgtgta ctacggcg 120
cccggtgtgg aaggaggccac caccacccctg ttctgcgcgc gcgcacgc 180
accgaggtgc acaaactgtgc ggcacccac gcctgcgtgc ccaccgcaccc caaccccccag 240
gagatgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catgggtggag 300
cagatgcacg aggacatcat cagcgtgtgg gaccagagcc tgaagccctg cgtgaagctg 360
accccccgtg gcgtggggcgc cggcaactgc cagaccagcg tgatcacca ggcctggccc 420
aagggtgagct tcgagcccat ccccatccac tactgcgcgc cgcgggctt cgcacatctg 480
aagtgcacg acaagaagtt caacggcgcg ggcctgtca ccaacgtgag caccgtgcag 540
tgcacccacg gcatccggccc cgtgggtgac acccagctgc tgctgaacgg cagcctggcc 600
gaggaggggcg tggatcccg cagcgagaac ttccacgc acgccaagac catcatctg 660
cagctgaagg agagcgtggc gatcaactgc accccggggca acaacaacac cgcgaagagc 720
atcaccatcg gccccggcccg cgccttctac gccacccggcg acatcatcg cgcacatccgc 780
caggcccaact gcaacatcg cggcgagaag tggaaacaaca ccctgaagca gatcggtacc 840
aagctcgagg cccagttcg cacaacgacc atcggttca agcagagcag cggcggccgc 900
cccgagatcg tgatgcacag cttcaactgc ggcggcgagt tcttctactg caacagcacc 960
cagctgttca acacgcacccgt gaacaacacc atccggggccca acaacacccaa cggcaccatc 1020
accctggccct gcccgcattca gcaacatcg aaccgtgtgc aggaggtggg caaggccatg 1080
tacgcccccc ccatccgggg ccagatccgc tgacgcgc acatcacccg cctgctgtcg 1140
acccgcgacg gcccgcacgg gatcgcaccc accacccgaga tcttccgcgc cggcggccgc 1200
gacatgcgcg acaactggcg cagcgagctg tacaagtata aggtgggtgaa gatcgagccc 1260
ctggggcggtgg ccccccaccaa ggcacgcgc cgcgtgggtgc agcgcgagaa ggcgcgcgtg 1320
accctggggcg ccatgttccct gggcttccgt ggcggccgc gcagcaccat gggcgccgc 1380
agcctgaccc tggatcccg cggccgcgcag ctgtgcgc gatcgtgcgc gcagcagaac 1440
aacctgtgc gcccgcattca gggccgcgcag caccgtgtgc agctgacccgt gtggggccatc 1500
aagcagctgc agggccgcgt gctggccgtg gagcgctacc tgaaggacca gcagctgtcg 1560
ggcatctggg gctgcagcgcc caagctgatc tgcacccaccc cctgccccctg gaacggccgc 1620
tggagcaaca agaggctggc ccagatctgg aacaacatcg cctggatggc 1680

gagatcgaca actacaccaa cctgatctac accctgatcg aggagagcca gaaccagcag 1740
 gagaagaacg agcaggagct gctggagctg gacaagtggg ccagectgtg gaactggtc 1800
 gacatcagca aytggctgtg gtacatcaag atcttcatca tgatcggtgg cggcctggtg 1860
 ggcctgcga tcgtgttac cgtgctgagc atcgtaacc gcgtgcgcca gggctacagc 1920
 cccctgagct tccagacccc cttecccccc ccccgccgccc ccgaccgccc cgagggcata 1980
 gaggaggagg gggcggagcg cgaccgcac ccgcagcagcc ccctgggtca cggcctgtgt 2040
 gccctgatct gggacgacct ggcagcctg tgcctgttca gctaccaccc cctgcgcgac 2100
 ctgatcctga tcgcccggcc catcggtgg ctgctggcc gccgcggctg ggaggccctg 2160
 aagtactggg gcaacctgtc gcagtaactgg atccaggagc tgaagaacag cgccgtgagc 2220
 ctgttcgacg ccatcgccat cgccgtggcc gagggcaccg accgcattcat cgaggtggcc 2280
 cagcgcatcg gccgcgcctt cctgcacatc cccgcgcga tccgcccaggg cttcgagcgc 2340
 gccctgctgt aactcgag 2358

<210> 51

<211> 1494

<212> DNA

<213> Human immunodeficiency virus

<400> 51

acaacagtct tgggttacccat agtctattat ggggtacctg tggaaaaga agcaaccacc 60
 actctgtttt gtgcatacaga tgctaaagca tacaaggcag aggcacataa cgtctgggt 120
 acacatgcct gtgtacccac agaccccaac ccacaggaaag taaatttaac aaatgtgaca 180
 gaaaattttt acatgtggaa aaataacatg gtggaaacaga tgcatgagga tataatcagt 240
 ttatgggatc aaagcctaaa gccatgtgtaa aatattaaccc cactctgtgt tactttaaat 300
 tgtactgata agttgacagg tagtactaat ggcacaaaataa gtacttagtgg cactaatagt 360
 actagtggca ctaatagtac tagtactaat agtactgata gttggaaaaa gatgccagaa 420
 ggagaaataaa aaaactgctc tttcaatatac accacaagtg taagagataa agtgcagaaa 480
 gaatattctc tcttctataaa acttgatgtaa gtaccaatag ataatgataa tgctagctat 540
 agattgataa attgtataaac ctcagtcatt acacaaggct gtccaaaggt atctttgaa 600
 ccaattcccc tacattattt tgccccggct ggttttgcga ttctaaagtg taaagataag 660
 aagttcaatg gaacaggacc atgtaaaaat gtacggcacag tacaatgcac acatggatt 720
 agaccatgt tatcaactca actgtgttta aatggcagtc tagcagaaga agagatagta 780
 cttagatctg aaaatttcac agacaatgct aaaaccataaa tagtacagct gaatgaatct 840
 gtagaaattt attgtataag acccaacaat aatacaagaa aaagtataca tataggacca 900
 gggagagcat tttatgcac accgtatata ataggagaca taagacaagc acattgtaac 960
 attagtaaattt caaaactggac taacacttta gaacagatag ttggaaaaattt aagagaacaa 1020
 ttgggatataaaat aatctttaat tcatttcac gaggggaccc agaaattgtt 1080
 ttccacagtt ttaattgtgg aggggattt ttctattgtt atacatcaca actatttaat 1140
 agtacctgga atattactga agaggtttt aagactaaag aaaatgacac tattcatactc 1200
 ccatgcacaa taagacaaat tataaacatg tggcaagaag tagggaaaagc aatgtatgcc 1260
 cttcccatca gaggacaaat taaatgttca tcaaatattt cagggctgtt attaactaga 1320
 gatgggtgtt ctaacaataa taggacaaac gacaccgaga ctttcagacc tgggggagga 1380
 aacatgaagg acaattggag aagtgaatta tataaatata aagtagtaag aattgaacca 1440
 ttaggatgtt caccacccca ggcaaaagaga agagtggtgc aaagagagaa aaga 1494

<210> 52

<211> 2007

<212> DNA

<213> Human immunodeficiency virus

<400> 52

acaacagtct tgggttacccat agtctattat ggggtacctg tggaaaaga agcaaccacc 60
 actctgtttt gtgcatacaga tgctaaagca tacaaggcag aggcacataa cgtctgggt 120
 acacatgcct gtgtacccac agaccccaac ccacaggaaag taaatttaac aaatgtgaca 180
 gaaaattttt acatgtggaa aaataacatg gtggaaacaga tgcatgagga tataatcagt 240
 ttatgggatc aaagcctaaa gccatgtgtaa aatattaaccc cactctgtgt tactttaaat 300
 tgtactgata agttgacagg tagtactaat ggcacaaaataa gtacttagtgg cactaatagt 360
 actagtggca ctaatagtac tagtactaat agtactgata gttggaaaaa gatgccagaa 420
 ggagaaataaa aaaactgctc tttcaatatac accacaagtg taagagataa agtgcagaaa 480

gaatatttctc	tcttctataa	acttgatgta	gtaccaatag	ataatgataa	tgcttagctat	540
agattgataa	attgtatatac	ctcagtcatt	acacaaggct	gtccaaaggt	atctttgaa	600
ccaattccca	tacattatttgc	tgccccggct	ggtttgcga	ttctaaagtgt	taaagataag	660
aagttcaatg	gaacaggacc	atgtaaaaat	gtcagcacag	tacaatgcac	acatgaaatt	720
agaccaggtag	tatcaactca	actgctgtt	aatggcagtc	tagcagaaga	agagatagta	780
cttagatctg	aaaatttcac	agacaatgct	aaaaccataa	tagtacagct	aatgaaatct	840
gtagaaaatta	attgtataag	acccaacaat	aatacaagaa	aaagtataca	tataggacca	900
gggagagcat	tttatgcaac	aggtgatata	ataggagaca	taagacaagc	acattgtaac	960
attagtaaag	caaactggac	taacacttta	gaacagatag	ttgaaaaattt	aagagaacaa	1020
tttgggataa	ataaaaacaat	aatctttaaat	tcatcctcag	gaggggaccc	agaaaattgta	1080
tttcacagtt	ttaattgtgg	aggggaattt	ttctattgtt	atacatcaca	actatttaat	1140
agtaccttgg	atattactga	agaggtaaat	aagactaaag	aaaatgacac	tatcataactc	1200
ccatgcagaa	taagacaaat	tataaacatg	tggcaagaag	tagaaaaagc	aatgtatgcc	1260
cctcccatca	gaggacaaat	taaatgttca	tcaaataattt	cagggctgc	attaactaga	1320
gatggtgtta	ctaacaataa	taggacgaac	gacaccgaga	ccttcagacc	tgggggagga	1380
aacatgaagg	acaattggag	aagtgaatta	tataaatata	aagttagtaag	aattgaacca	1440
tttaggagtag	cacccaccca	ggcaaagaga	agagtggtgc	aaagagagaa	aagagcagtg	1500
ggacttaggg	ctttgttcat	tgggttcttg	ggagcagcag	gaagcactat	gggcgcagcg	1560
tcagtgacgc	tgacggtaca	ggccagacaa	ttattgtctg	gtatagtgc	acagcagaac	1620
aatttgcgt	gagtttattga	ggcgcacacag	catctgttgc	aactcacgg	ctggggcatc	1680
aaacagctcc	aggcaagaat	cctggctgtg	gaaagatacc	taaaggatca	acagctctta	1740
gggattttggg	gttgctctgg	aaaactcatt	tgcaccacta	ctgtgccttg	gaactctagt	1800
tggagtaata	aatctctgac	tgagattttgg	gataatatga	cctggatgga	gtggaaaaga	1860
gaaattggca	attatacagg	cttaatatac	aatttaatttgc	aaatagcaca	aaaccagcaa	1920
gaaaagaatg	aacaagaatt	atttggatata	gacaagtggg	caagttgttgc	gaattgttt	1980
gatataacaa	actggctgttgc	gtatata				2007

<210> 53
<211> 2532
<212> DNA
<213> Human immunodeficiency virus

```

<400> 53
acaacagtct tgtgggtcac agtctattat ggggtacctg tggaaaga agcaaccacc 60
actctgttt gtcatcaga tgctaaagca tacaaagcg aggacataa cgtctggct 120
acacatgcct gtgtacccac agaccccaa ac ccacaggaag taaatcca aatgtgaca 180
aaaattta acatgtggaa aaataacatg gtggacaga tgcatgagga tataatcagt 240
ttatggatc aaaggctaaa gccatgtgt aaattaaccc cactctgtgt tactttaaat 300
tgtactgata agttgacagg tagtactaat ggcacaaata gtactagtgg cactaatagt 360
actagtggca ctaatagttac tagtactaat agtactgata gttggaaaa gatgccagaa 420
ggagaaaataa aaaactgctc ttcaatatac accacaagtg taagagataa agtgcagaaa 480
gaatattctc tcttcataa acttgatgta gtaccaatag ataatgataa tgctagctat 540
agattgataa attgtataac ctctgtt acacaaggct gtccaaagggt atctttgaa 600
ccaattccca tacattattg tgccccggct ggtttgcga ttctaaagtg taaagataag 660
aagttcaatg gaacaggacc atgtaaaaat gtcagcacag tacaatgcac acatgaaatt 720
agaccagtag tatcaactca actgctgtt aatggcagtc tagcagaaga agagatagta 780
cttagatctg aaaatttcac agacaatgct aaaaccataa tagtacagct gaatgaatct 840
gtagaaaatta attgtataag acccaacaat aatacaagaa aaagtataaca tataggacca 900
gggagagcat ttatgcaac aggtgatata ataggagaca taagacaagg acattgtaac 960
attagtaaag caaactggac taacacttta gaacagatag ttggaaaatt aagagaacaa 1020
tttgggaaata ataaaacaat aatctttaaat tcatttcag gaggggaccc agaaattgt 1080
tttcacagtt ttaattgtgg aggggaattt ttctattgtt atacatcaca actatttaat 1140
agtacctgga atattactga agaggtaaat aagactaaag aaaatgacac tatcataactc 1200
ccatgcagaa taagacaaat tataaacatg tggcaagaag tagggaaaagg aatgtatgcc 1260
cctcccatca gaggacaaat taaatgttca tcaaataat caggctgtt attaactaga 1320
gatgggtggtt ctaacaataa taggacgaac gacaccgaga ctttcagacc tgggggagga 1380
aacatgaagg acaattggag aagtgaatta tataaatata aagttagtaag aattgaacca 1440
ttaggagtag cacccaccca ggcacaaagaga agagttgtgc aaagagagaa aagagcagt 1500
ggactaggag ctttggatcat tgggttcttgg ggcacggcagc gaagcactat gggcgcagcg 1560

```

tcagtgcacgc	tgacgggtaca	ggccagacaa	ttatttgtctg	gtatagtgc	acagcagaac	1620
aatttgcgtga	gagctattga	ggcgcaacag	catctgttc	aactcacgg	ctggggcatc	1680
aaacagctcc	aggcaagaat	cctggctgt	gaaagatacc	taaaggatca	acagctcta	1740
gggattttggg	gttgctctgg	aaaactcatt	tgcaccacta	ctgtgcctt	gaactctagt	1800
tggagtaata	aatctctgac	tgagattttgg	gataatatga	cctggatgg	gtgggaaaga	1860
gaaattggca	attatacagg	cttaatatac	aatttaattt	aaatagcaca	aaaccagcaa	1920
gaaaagaatg	aacaagaatt	attggaaatta	gacaagtggg	caagttgt	gaattgttt	1980
gatataacaa	actggctgt	gtatataaga	atattcataa	tgatagttag	aggcttgata	2040
ggtttaagaa	tagttttgc	tgtactttct	atagtgaaa	gagtttaggc	gggataactca	2100
ccaatattcat	tgcagaccccg	cctcccagct	cagaggggc	ccgacaggcc	cgaaggaatc	2160
gaagaagaag	gtggagagag	agacagagac	agatccaatc	gattagtgc	tggattattt	2220
gcactcatct	gggacgatct	gcggagcc	tgccttca	gctaccaccc	cttgagagac	2280
ttactcttga	ttgtagcgag	gatttgtggaa	cttctggac	gcaggggggt	ggaagccctc	2340
aagtatttggt	ggaatctcct	gcagtatttt	agtcaggagc	taaagagat	tgctgttagt	2400
ttgttttaatg	ccacagcaat	agcagtagct	gaagggacag	ataggattat	agaaaatagta	2460
caaagaattt	ttagagctgt	aattcacata	cctagaagaa	taagacagg	cttggagagg	2520
gctttactat	aa					2532

<210> 54
<211> 1599
<212> DNA
<213> Arti

<220>
<223> Description of Artificial Sequence: qp120.modUS4

<400> 54
gaattcggcca ccatggatgc aatgaagaga gggctctgct gtgtgtcgct gctgtgtgga 60
gcagtcttcg tttcgccccag cgccaccacc gtcgtgtgg tgaccgtgtta ctacggcg 120
cccgtgtgga aggaggccac caccacccctg ttctgcggca ggcacgccaa ggcttacaag 180
ggcgaggccc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caaccccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgag 300
cagatgcattg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
accccccgtg gcgtgaccct gaactgcacc gacaagctga cccggcagcac caacggcacc 420
aacagcacca gcgccaccaa cagcaccagc ggcaccaaca gcaccagcac caacagcacc 480
gacagctggg agaagatgcc cgagggcgag atcaagaact gcagcttcaa catcaccacc 540
agcgtgcgcg acaagggtgca gaaggaggta acgcctgttca acaagctgga cgtggtggcc 600
atcgacaaacg acaacgcccag ctaccgcctg atcaactgca acaccagcgt gatcaccagg 660
gcctgccccca aggtgagctt cgagcccatc cccatccact actgcggcccc cgccggcttc 720
gccatcctga agtgcacagga caagaagttt aacggcaccg gcccctgcaaa gaacgtgagc 780
acccgtgcagt gcacccacgg catccgcctt gttggtagca cccagctgct gctgaacggc 840
agcctggccg aggaggagat cgtgtcgcc tcggagaact tcaccgacaa cgccaagacc 900
atcatcgtgc agctgaacga gtccgtggag atcaactgca tcggcccaa caacaacacg 960
cgtaagagca tccacatcggtt ccccgccgc gccttctacg ccaccggcga catcatcg 1020
gacatccggcc aggcccactg caacatcagc aaggccaact ggaccaacac cctcgagcag 1080
atcggtggaga agctgcgcga gcagttcgcc aacaacaaga ccatcatctt caacagcagc 1140
agcggcggcg accccggagat cgtgttccac agcttcaact gggcgccga gttttctac 1200
tgcaacacca gccagctgtt caacagcacc tggaaatca ccggaggaggt gaacaagacc 1260
aaggagaacg acaccatcat cctgccttgc cgcatccggc agatcatcaa catgtggcag 1320
gaggtggcca aggccatgtt cggccccc atccggggcc agatcaagtg cagcagcaat 1380
attaccggcc tgcgtgtgac cggcgacggc ggcaccaaca acaaccgcac caacgcacacc 1440
gagaccttcc gccccggccgg cggcaacatg aaggacaact ggcgcagcga gctgtacaag 1500
tacaagggtgg tgcgcacatcgat gccccctggc gtggccccc cccaggccaa ggcggcg 1560
gtgcagcgcg agaagcgcta agatatecgta tcctcttaga 1599

<210> 55
<211> 1350
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp120.modUS4.del 128-194

<400> 55

```

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgctgct gctgtgtgga 60
gcagtctcg tttcgcccaag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccacccctg ttctgcccga gcgacgcaca ggcttacaag 180
gccgaggccc acaacgtgtg ggcacccac gcctgctgtc ccaccgaccc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtggag 300
cagatgcgtg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
accccccctgt gcgtgggggc aggaaactgc gagaccagcg tgatcacca ggcctgcccc 420
aaggtagct tcgagcccat cccatccac tactgccc cgcgggctt cgccatccctg 480
aagtgcagg acaagaagtt caacggcacc ggcctgtca agaacgtgag caccgtgcag 540
tgcacccacg gcatccgccc cgtggtagcc acccagctgc tgctgaacgg cagctggcc 600
gaggaggaga tcgtgctgca ctccgagaac ttcaacggaca acgccaagac catcatcg 660
cagctgaacg agtccgtgg aatcaactgc atccggccca acaacaacac gcttaagagc 720
atccacatcg gccccggccg cgcccttctac gccaccggcg acatcatcg cgacatccgc 780
caggccact gcaacatcg caaggccaaac tggaccaaca ccctcgagca gatcgtagg 840
aagctgcgcg agcagttcg caacaacaag accatcatct tcaacagcag cagccggccg 900
gaccccgaga tcgtgttcca cagcttcaac tgccggccg agttcttca ctgcacacacc 960
agccagctgt tcaacagcac ctggAACATC accggaggagg tgaacaagac caaggagaac 1020
gacaccatca ttctgcccctg ccgcattccgc cagatcatca acatgtgca ggaggtggc 1080
aaggccatgt acgccccccc catccggcc cagatcaagt gcagcagcaa tattaccggc 1140
ctgctgtca cccggcggccg cgccaccaac aacaaccgca ccaacgacac cgagaccc 1200
cgccccggccg gccggcaacat gaaggacaac tggcgeagcg agctgtacaa gtacaagg 1260
gtgcgcacatcg agcccccggg cgtggccccc acccaggcca agcgccgct ggtcagcgc 1320
gagaagcgct aagatatcg atcctctaga                                1350

```

<210> 56

<211> 2112

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: gp140.modUS4

<400> 56

```

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgctgct gctgtgtgga 60
gcagtctcg tttcgcccaag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccacccctg ttctgcccga gcgacgcaca ggcttacaag 180
gccgaggccc acaacgtgtg ggcacccac gcctgctgtc ccaccgaccc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtggag 300
cagatgcgtg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
accccccctgt gcgtgggggc aggaaactgc gagaccagcg tgatcacca ggcctgcccc 420
aacagcacca gccggcacca cagcaccacc ggcaccaaca gcaccagcac caacagcacc 480
gacagctggg agaagatgcc cgaggccgag atcaagaact gcagcttca catcaccacc 540
agcgtgcgcg acaagggtca gaaggagatc agcctgttca acaagctgca cgtgtggcc 600
atcgacaacg acaacgcaccc ctaccggctg atcaactgca acaccagcgt gatcaccacc 660
gcctggccca aggtgagctt cgagccatc cccatccact actgccc cgcggcc 720
cccatcttca agtgcaaggg caagaagttt aacggcaccc gcccctgcaaa gaacgtgagc 780
accgtgcagt gcaacccacgg catccggcc gtcgtggagca cccagctgt gctgaacggc 840
agcctggccg aggaggatc cgtgtgcgc tccgagaact tcaccgacaa cgccaaagacc 900
atcatcgatcg agctgaacgca gtccgtggag atcaactgca tccggcccaaa caacaacacg 960
cgtaagagca tccacatcg ccccgccgc gccttctac ccaccggcga catcatcg 1020
gacatccgccc agggccactg caacatcg aaggccaaact ggaccaacac cctcgagcag 1080
atcggtggaga agctgcgcga cgatccgca aacaacaaga ccatacatctt caacagcagc 1140
agcggccggccg acccccgagat cgtgtttcac agcttcaact gcccggccgaa gttcttctac 1200
tgcaacacca gccagctgtt caacagcaccc tggaaacatca ccgaggaggt gaacaagacc 1260

```

aaggagaacg acaccatcat cctgcctgc cgcatccgccc agatcatcaa catgtggcag 1320
 gaggtggca aggccatgta cgcccccccc atccgcggcc agatcaagt cagcagcaat 1380
 attaccggcc tgctgctgac ccgcgacggc ggcaccaaca acaaccgcac caacgacacc 1440
 gagaccttcc gccccggcg ggcaacatg aaggacaact ggccgagcga gctgtacaag 1500
 tacaagggtgg tgcgcatcg gccccctggc gtggggggca cccaggccaa ggcggcgctg 1560
 gtgcagcgcg aagaagcgccg cgtggcctg ggcccctgt tcatcggtt cctggggcgcc 1620
 gccgggagca ccatgggcgc cgcctccgtg accctgaccg tgcaggcccg ccagctgctg 1680
 agcggcatcg tgcagcagca gaacaacctg ctgcgcgcga tcgaggccca gcagcacctg 1740
 ctgcagctga ccgtgtgggg catcaagcag ctgcaggccc gcatacctggc cgtggagcgc 1800
 tacctgaagg accagcagct gctggcatt tggggctgca gcccgaagct gatctgcacc 1860
 accaccgtgc cctggaacag cagctggagc aacaagagcc tgaccgagat ctgggacaac 1920
 atgacctgga tggagtggga ggcggagatc ggcactaca cggcctgt ctacaacactg 1980
 atcgagatcg cccagaacca gcaggagaag aacgagcagg agctgctgga gctggacaag 2040
 tggggcagcc tggactg gttcgacatc accaactggc tgggtacat ctaagatatc 2100
 ggatcctcta ga 2112

<210> 57

<211> 2112

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.mut.modUS4

<400> 57

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
 gcagtcttcg tttcgcccaag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgctg 120
 cccgtgtgga aggaggccac caccacccctg ttctgcgcga ggcacgcacaa ggcttacaag 180
 gcccaggccc acaaactgtg ggcacccac gcctgcgtgc ccacccgaccc caaccccccag 240
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgtt ggaagaacaa catgtggag 300
 cagatgcattg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 accccctgt gcgtgaccctt gaaactgcacc gacaagctga cccgcgcac caacggcacc 420
 aacagcacca gccggcacca cagcaccacc ggcaccaaca gcaccacgc acacagcacc 480
 gacagtggg agaagatgcc cgaggccgag atcaagaact gcagcttcaa catcaccacc 540
 agcgtgcgcg acaagggtgca gaaggagatc agcctgttctt acaagctgga cgtgggtgccc 600
 atcgacaacg acaacgcacc ctaccgcctg atcaactgca acaccagcgt gatcaccacc 660
 gcctggccca aggtgagctt cgagccatc cccatccact actgcggcccc cgccggcttc 720
 gccatcttga agtgcaagga caagaagtcc aacggcaccg gcccctgcaaa gaacgtgagc 780
 accgtgcagt gcacccacgg catccggccc gtgtgagca cccagctgt gctgaacggc 840
 agcctggccg aggaggatg cgtgctgcgc tccgagaact tcaccgacaa cgccaaagacc 900
 atcatctgtgc agtgcacca gtcctgtggg atcaactgca tccggcccaaa caacaacacg 960
 cgtaagagca tccacatcggtt ccccgccgc gccttctacg ccacccggcga catcatcgcc 1020
 gacatccgcg aggcccactg caacatcgac aaggccaaact ggaccaacac cctcgagcag 1080
 atcggtggaga agtgcgcgca gcagttcgcc aacaacaaga ccacatctt caacagcagc 1140
 agcggccgcg acccccgatg cgtgtttcac agcttcaact gccggccgcga gttcttctac 1200
 tgcaacacca gccagctgtt caacagcacc tggaaacatca ccgaggaggt gaacaagacc 1260
 aaggagaacg acaccatcat cctgcctgc cgcacccgc accatccgcgca agatcatcaa catgtggcag 1320
 gaggtggca aggccatgtt gcccccccccc atccgcggcc agatcaagt cagcagcaat 1380
 attaccggcc tgcgtgtgac ccgcgcacccgg ggcaccaaca acaaccgcac caacgacacc 1440
 gagaccctcc gccccggcg ggcaacatg aaggacaact ggccgagcga gctgtacaag 1500
 tacaagggtgg tgcgcatcg gccccctggc gtggggggca cccaggccaa ggcggcgctg 1560
 gtgcagcgcg agaagagcgc cgtggccctg ggcggccctgt tcatcggtt cctggggcgcc 1620
 gcccggagca ccacccggcc cgcctccgtg accctgaccg tgcaggcccg ccagctgctg 1680
 agcggcatcg tgcagcagca gaacaacctg ctgcgcgcga tcgaggccca gcagcacctg 1740
 ctgcagctga ccgtgtgggg catcaagcag ctgcaggccc gcatacctggc cgtggagcgc 1800
 tacctgaagg accagcagct gctggcatt tggggctgca gcccgaagct gatctgcacc 1860
 accaccgtgc cctggaacag cagctggagc aacaagagcc tgaccgagat ctgggacaac 1920
 atgacctgga tggagtggga ggcgcgatc ggcaactaca cggcctgtat ctacaacactg 1980

atcgagatcg cccagaacca gcaggagaag aacgagcagg agctgctgga gctggacaag 2040
 tggccagcc tgtgaaactg gttcgacatc accaactggc tgtggtacat ctaagatatac 2100
 ggatccctcta ga 2112

<210> 58
 <211> 2181
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: gp140TM.modUS4

<400> 58

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgctgct gctgtgtgga 60
 gcagtcttcg tttcgccccag cgccaccacc gtgctgtggg tgaccgtgtta ctacggcgtg 120
 cccgtgtgga aggaggccac caccacccctg ttctgcccga gcgacgccaa ggcttacaag 180
 gcccaggccc acaacgtgtt ggccacccac gcctgcgtgc ccacccgaccc caaccccccag 240
 gaggtgaaacc tgaccaacgtt gaccgagaac ttcaacatgtt ggaagaacaa catggtgag 300
 cagatgcattt aggacatcat cagcctgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
 acccccctgt gcgtgaccctt gaaactgcacc gacaagctgtt ccggcagcac caacggcacc 420
 aacagcacca gcgccaccaa cagcaccacgg ggccaccaaca gcacccagcac caacagcacc 480
 gacagctggg agaagatgcc cgagggcgag atcaagaactt gcagcttcaa catcaccacc 540
 agcgtgcgcg acaagggttca gaaggagttt agcctgttca acaagctgtt cgtgtgccc 600
 atcgacaaacg acaacgcggc ctaccgcctt atcaactgttca acaccaggctt gatcaccacc 660
 gcctgccccca aggtgagctt cgagccatc cccatccactt actgcgcggcc cgccggcttc 720
 gccatccttga agtgcaagggaa caagaagttt aacggccacccg gcccctgcaaa gaacgtgagc 780
 accgtgcagt gcacccacggg catccggccccgtt gtgttggatca cccagctgtt cgtgaacggc 840
 agcctggcccg aggaggatgtt cgtgtgcgc tccggaaactt tcaccggacaa cgccaaagacc 900
 atcatctgttca agtgcacca gttccgtgggatcaactgttca tccggcccaaa caacaacacg 960
 cgtaagagca tccacatcggg ccccgccgc gccttctacgg ccacccggcga catcatcgcc 1020
 gacatccgcg aggcccacttga caacatccggc aaggccaaactt ggaccaacac cctcgagcag 1080
 atcggtggaga agtgcgcgcgca gcagtttgcgc aacaacaaga ccattcatctt caacagcagc 1140
 agcggccggcg accccggatgtt cgtgtttccactt agtgcaccaactt gccggccggcga gttcttctac 1200
 tgcaacacca gccagctgtt caacagcacc tggaaacatca ccggaggaggtt gaacaagacc 1260
 aaggagaacg acaccatcat cctgccttgc cgcattccggcc agatcatctt catgtggcag 1320
 gaggtggggca aggccatgttca cggccggccatccggcc agatcaagtgtt cagcagcaat 1380
 attaccggcc tgcgtgttgc cccgcacccggc ggcaccaaca acaacccgcac caacgacacc 1440
 gagaccccttcc gccccggccgg cggcaacatgtt aaggacaactt ggccggccggcga gctgtacaag 1500
 tacaagggtgg tgcgcacatca gccccctggc gtggccccc cccagggccaa ggcggccgtg 1560
 gtgcagccgcg agaaggccgcg cgtggccctgtt ggcggccctgtt tcattggctt cctggccggcc 1620
 gcccggggccca ccatggccgcg cccctccgtt accctgaccgtt tgccggcccg ccagctgtgt 1680
 agcggccatcg tgcagcagca gaacaaccttgc tgcgcgcgc tgcggccca gcagcacctt 1740
 ctgcagctgttccgtgtggg catcaaggccatcg tgcaggccgc gcattccgtt cgtggagccgc 1800
 tacctgttggacc accagccatgttgc tggggctgtca gccggcaagttt gatctgcacc 1860
 accaccgttgc cctggaaacatgttca cagctggggatca aacaagagcc ttgaccggatgtt ctgggacaac 1920
 atgacccatgttca tggagttgggatca ggcggccatca cccggccctgtt ctacaaccttgc 1980
 atcgagatcg cccagaacca gcaggagaag aacgagcagg agctgtgttca gctggacaag 2040
 tggccagcc tggggacttgc gttcgacatc accaactggc tgggttacat ccgcatttc 2100
 atcatatgttca tggccggccatcg gatccggccatcg cgcattgttgc tgcggccgtt gaggcatcg 2160
 taagatatacg gatcccttagt a 2181

<210> 59
 <211> 1818
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 gp140.modUS4.delV1/V2

<400> 59

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
 gcagtctcg tttcgcccaag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
 cccgtgtgga aggaggccac caccacccctg ttctgcgcca gcgacgcca ggcttacaag 180
 gccgaggccc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caacccccag 240
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
 cagatgcattg aggacatcat cagcctgtgg gaccagagcc tgaaggccctg cgtggggcc 360
 ggcaggcct gccccaaagggt gagcttegag cccatccccca tccactactg cgccccggcc 420
 ggctcgcca ttctgaagtg caaggacaag aagttaacg gcacccggcc ctgcaagaac 480
 gtgagcaccg tgcagtgcac ccacggcatttgcgccccctg tgagcacca gctgctgctg 540
 aacggcagcc tggccgagga ggagatcgatg ctgcgcctccg agaacttcac cgacaacgccc 600
 aagaccatca tcgtgcagct gaacgagtcc gtggagatca actgcattccg ccccaacaac 660
 aacacgcgtt agagcatcca catcgcccc gggcgcgcct tctacgcac ccggcgcacatc 720
 atcggcaca tccgcccaggc ccactgcaac atcagcaagg ccaactggac caacaccctc 780
 gagcagatcg tggagaagct ggcgcgacatg ttcggcaaca acaagaccat catcttcaac 840
 agcagcagcg gggcgcaccc cgagatcgatg ttccacagct tcaactgcgg cgccgagttc 900
 ttctactgca acaccagcca gctgttcaac agcacctgga acatcacca ggaggtgaac 960
 aagaccaagg agaacgcacatc catcattctg cccgcgcac tccgcccagat catcaacatg 1020
 tggcaggagg tggcaaggc catgtacgc ccccccattcc gcccgcagat caagtgcagc 1080
 agcaatatta cccgcctgt gctgacccgc gacggcggca ccaacaacaa ccgcaccaac 1140
 gacaccgaga cttccgcggc cggcggcggc aacatgaagg acaactggcg cagcgagctg 1200
 tacaagtaca aggtggtgcg catcgaccc ctggcgctgg ccccccacca ggcacagcgc 1260
 cgcgtggcgc agcgcgagaa ggcgcgcgtg ggcctggcg ccctgttcat cgggttcctg 1320
 ggcgcgcgcg ggagcaccat gggcgcgcgc tccgtgaccc tgaccgtca ggcgcgcgc 1380
 ctgctgagcg gcatcgatgcg acagcagaac aacctgtgc ggcgcacatcg ggcgcacagc 1440
 cacctgtgc agctgaccgt gtggggcatttgc aagcagctgc agggccgcatttgc cctggcgctg 1500
 gagcgctacc tgaaggacca gcaagctgtg ggcattctggg gctgcagcgg caagctgatc 1560
 tgcaccacca ccgtgcctcg gaacagcgc tggagcaaca agagcctgac cgagatctgg 1620
 gacaacatga cctggatgga gtgggagcgc gagatcgacca actacaccgg cctgatctac 1680
 aacctgtatcg agatcgccca gaaccagcgc gagaagaacg agcaggagct gctggagctg 1740
 gacaagtggg ccagcctgtg gaactggatc gacatcacca actggctgtg gtacatctaa 1800
 gatatcgat cctctaga 1818

<210> 60

<211> 2031

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.modUS4.delV2

<400> 60

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
 gcagtctcg tttcgcccaag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
 cccgtgtgga aggaggccac caccacccctg ttctgcgcca gcgacgcca ggcttacaag 180
 gccgaggccc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caacccccag 240
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
 cagatgcattg aggacatcat cagcctgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
 accccctgt gctgtacccttgc gaaactgcacca gcaacgcgttgc ccggcgcac caacggcacc 420
 aacagcacca gcccgcacca cagcaccac ggcaccaaca gcacccgcac caacagcacc 480
 gagcgtggg agaagatgcg cgaggcgcgcg atcaagaacatc gcaacttcaa catcgccgc 540
 ggcgcctgt tcaactgcacca caccagcgatg atcaccacgg cctggccca ggtgagctc 600
 gagcccatcc ccatccacta ctgcgcggcc gcccggcttc ccatecttgcgttgc gtcacaggac 660
 aagaagtta acggcaccgg cccctgtcaag aacgttgacca cctgtgcgttgc caccacggc 720
 atccggcccg tgggtgacccatc ccagctgtgc ctgaacggcc gctggccca ggaggagatc 780
 gtgctgcgtt cccgagaactt caccgacaac gccaagacca tcaactgcgttgc gctgaacgc 840
 tccgtggaga tcaactgcatttgc cccggccac aacaaacacgc gtaagagcat ccacatcgcc 900
 cccggccgcg cttctacgc caccggcgcac atcatcgccca ggcgcacttgc 960

aacatcagca aggccaactg gaccaacacc ctcgagcaga tcgtggagaa gctgcgcgag 1020
 cagttcgca acaacaagac catcatctt aacagcagca gcggcggcga ccccgagatc 1080
 gtgttccaca gcttcaactg cgccggcgag ttcttctact gcaacaccag ccagctgttc 1140
 aacagcacct ggaacatcac cgaggaggtg aacaagacca aggagaacga caccatcatc 1200
 ctgccctgca gcatccgcca gatcatcaac atgtggcagg aggtggccaa ggccatgtac 1260
 gcccccccca tcccgccca gatcaagtgc aacagcaata ttaccggccct gctgctgacc 1320
 cgcgacggcg gcaacaaaca caaccgcacc aacgacaccg agaccttccg ccccgccg 1380
 ggcaacatga aggacaactg ggcgcggag ctgtacaagt acaagggttgt ggcgcacatcg 1440
 cccctggcg tgccccccac ccaggccaag cgcgcgtgg tgcagcgcga gaagcgcg 1500
 gtggggctgg ggcgcctgtt catcggttc ctggggcgcc cgccggagcac catgggcg 1560
 gcctccgtga ccctgacgt gcaggcccgc cagctgtga gccgcategt gcagcagcag 1620
 aacaacctgc tgcgcgcatt cgaggcccaag cagcacctgc tgcagctgac cgtgtggggc 1680
 atcaaggcgc tgcaggcccg catccctggcc gtggagcgtt acctgaaggaa ccagcagctg 1740
 ctgggcattt gggctgcag cggcaagctg atctgcacca ccaccgtgcc ctggAACAGC 1800
 agctggagca acaagagcc gaccgagatc tggacaaca tgacctggat ggagtgggag 1860
 cgcgagatcg gcaactacac cggcctgtatc tacaacctga tgcagatcgc ccagaaccag 1920
 caggagaaga acgagcagga gctgctggag ctggacaagt gggccagct gtggaaactgg 1980
 ttcgacatca ccaactggct gtggatatacg gatcccttag a 2031

<210> 61

<211> 1818

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.mut.modUS4.delV1/V2

<400> 61

gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgctgtc gctgtgtgg 60
 gcagtttcg ttccgcccag cgccaccacc gtgtgtggg tgaccgtgtt ctacggcg 120
 cccgtgtgga aggaggccac caccacccgt ttctgcgcac ggcacgcacaa ggcttacaag 180
 gcccaggccc acaacgtgtg ggcacccac gcctgcgtgc ccaccgaccc caacccccc 240
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtg 300
 cagatgcattt aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtggccg 360
 ggcctggccct gcccccaaggt gagcttcgag cccatccccca tccactactg cgccccc 420
 ggcttcgcca tccctgaagtg caaggacaag aagtcaacg gcaccggccc ctgcaagaac 480
 gtgagcaccg tgcagtgcac ccacggcattt cgcccccgtgg tgagcaccca gctgctgt 540
 aacggcagcc tggccgagga ggagatctgt ctgcgcctccg agaacttcaac cgacaacg 600
 aagaccatca tggcgcagct gaacgagtcc gtggagatca actgcattcg ccccaacaac 660
 aacacgcgtt agagcatcca catcgcccccc ggcgcgcct tctacgcaccc cggegacatc 720
 atcggcgaca tccggcaggc ccactgcaac atcagcaagg ccaactggac caacaccctc 780
 gagcagatcg tggagaagct ggcgcgcgtt ttccgcacca acaagaccat catcttcaac 840
 agcagcagcg gccgcgcaccc cgagatctgt ttccacagct tcaactgcgg cggcgagttc 900
 ttctactgtt acaccagccca gctgttcaac agcacctggaa acatcaccga ggaggtgaac 960
 aagaccaagg agaacgcacac catcatcttgc ccctgcccgc tccgcgcattt catcaacatg 1020
 tggcaggagg tggcaaggc catgtacgc ccccccattt gccgcgcattt caagtgcagc 1080
 agcaatatta cccggctgtt gctgacccgc gacggcgccca ccaacaacaa ccgcaccaac 1140
 gacaccgaga cttccgcggcc cggcgccggc aacatgaagg acaactggcg cagcgagctg 1200
 tacaagtata aggtggtgcg catcgaccc ctggcggtgg ccccccaccca ggccaaagcgc 1260
 cgcgtgtgc agcgcgcgaa gacgcgcgtt ggcgcgttgc ccctgttcat cggcttcctg 1320
 ggcgcgcgcg ggagccatc gggcgccgc tccgtgaccc tgaccgtgtca ggccgcgc 1380
 ctgctgtgcg gcatgtgc gcaacgcaccc aacctgtgc ggcgcacccca ggcccgac 1440
 cacctgtgc agctgaccgt gtggggcattt aacgcacgtgc aggcggccat cctggccgt 1500
 gagcgttacc tgaaggacca gcaacgtgtt ggcacatctgg gctgcagcgg caagctgtatc 1560
 tgcaccacca cctgtgcctt gacacgcaccc tggaccaaca agacgcgttgcg cagatctgg 1620
 gacaacatga cttggatggaa gtggggcgcgc gacatcgccca actacaccgg cctgtatctac 1680
 aacctgtatcg agatcgccca gacccgcaccc gagaagaacg agcaggagct gctggagctg 1740
 gacaagtgggg ccagcctgtt gacatgttgc gacatcacca actggctgtt gatcatctaa 1800

gatatacgat cctctaga

1818

<210> 62

<211> 1818

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.modUS4.del 128-194

<400> 62

gaattcggca ccatggatgc aatgaagaga gggctctgt gtgtgctgt gctgtgtgga 60
 gcagtcttcg tttcgccca cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
 cccgtgtgga aggaggccac caccacccctg ttctgegccca ggcacgcca ggcttacaag 180
 gcccaggccc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caacccccag 240
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
 cagatgcatttgggaggcatc cagcctgtgg gaccagagcc tgaagccctg cgtgggcggcc 360
 ggcaggccct gcccccaaggt gagcttcgag cccatccccca tccactactg cgcggccggcc 420
 ggcttcggcca tcctgaagtg caaggacaag aagtcaacg gcacccggcccttgcagaac 480
 gtgagcaccc tgcaagtgcac ccacggcatttgcgtgg tgagcacccca gctgtgtg 540
 aacggcagcc tggccgagga ggagatcgatc ctgcgtcccg agaacttcac cgacaacgccc 600
 aagaccatca tcgtgcagcttcaacg aacgcgtcc ttccatcccg ccccaacaac 660
 aacacgcgttcaacgatccatc catcgccccc ggccggcccttgcaccccgccggcagatc 720
 atcggcgaca tccggccaggccactgtcaac atcagcaagg ccaactggac caacaccctc 780
 gagcagatcg tggagaagct ggcgcggccatc ttccggcaaca acaagaccat catttcaac 840
 agcagcagcccgccggccatc cgagatcgatc ttccacagct tcaactgcgg cggcggatcc 900
 ttctactgca acaccagccatc gctgttcaac agcacctggatc acatcaccga ggaggtgaac 960
 aagaccaagg agaacgcacatc catcatctgttccggccatc tccggccatcatcaacatg 1020
 tggcaggagg tggcaaggccatgtacggcc ccccccattcc gcccggccatcatc caagtgcagc 1080
 agcaatatta cccggccatc gctgaccggccatc gacggccggca ccaacaacaa ccgcaccaac 1140
 gacaccgaga cttccggccatc cggccggccatc aacatgaagg acaactggccatc cagcggatcc 1200
 tacaagtaca aggtggatcgatc catcgccccc ctggccgtgg ccccccacca ggccaaaggcc 1260
 cgcgtgtgc agcgcgagaa gagcgcgtg ggccctggccatc cccgttccat cggccatcc 1320
 ggcgcggccatc ggagcaccatc gggccggccatc tccgtgaccc tgaccgtgca ggccggccatc 1380
 ctgctgatcgatc gcatcgatc gcaatcgatc aacatcgatc ggcgcgttccatc ggcccgccatc 1440
 cacatcgatc agctgaccgttccatc aagcagatcgatc aggcggccatc cctggccgtg 1500
 gagcgttacc tgaaggatcgatc gcaatcgatc ggcgcgttccatc gtcgcggccatc 1560
 tgcaccatcgatc cctggccatc gacgcgttccatc tggagcaaca aggcgcgttccatc 1620
 gacaacatcgatc cctggatcgatc gtggggccatc gagatcgatc actacacccggccatc 1680
 aacatcgatcgatc agatcgatcgatc gacgcgttccatc gcaatcgatc gtcgcggccatc 1740
 gacaagtggg ccaggccatc gacgcgttccatc gacatcgatc actggccatc gtcgcggccatc 1800
 gatatacgat cctctaga 1818

<210> 63

<211> 1863

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp140.mut.modUS4.del 128-194

<400> 63

gaattcggca ccatggatgc aatgaagaga gggctctgt gtgtgctgt gctgtgtgga 60
 gcagtcttcg tttcgccca cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
 cccgtgtgga aggaggccac caccacccctg ttctgegccca ggcacgcca ggcttacaag 180
 gcccaggccc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caacccccag 240
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300

cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 accccccctgt gcgtgggggc agggaaactgc gagaccagcg tgatcaccca ggcctgcccc 420
 aaggtagct tcgagccat ccccatccac tactgcgccc cgcgggctt cgccatcctg 480
 aagtgcagg acaagaagtt caacggcacc ggccctgtca agaacgttag caccgtgcag 540
 tgcacccacg gcatccgccc cgtggtgagc acccagctgc tgctgaacgg cagctggcc 600
 gaggaggaga tcgtgctcg ctccgagaa acccatcgatac acgccaagac catcatcg 660
 cagctgaacg agtccgtgga gatcaactgc atccggccca acaacaacac gctgaagagc 720
 atccacatcg gccccggccg cgccttctac gccaccggcg acatcatcg cgacatccgc 780
 caggcccact gcaacatca gcaaggccaa tggaccaaca ccctcgagca gatcggtggag 840
 aagctgcgcg agcagttcg caacaacaag acccatcatct tcaacagcag cagccggcgc 900
 gaccccgaga tcgtgttcca cagcttcaac tgccggggcg agttcttta ctgcaacacc 960
 agccagctgt tcaacagcac ctggaacatc accggaggagg tgaacaagac caaggagaac 1020
 gacaccatca tcctgcccctg cgcacatccgc cagatcatca acatgtggca ggaggtgggc 1080
 aaggccatgt acggccccc catccgcggc cagatcaagt gcagcagcaa tattaccggc 1140
 ctgctctga cccgcgacgg cggcaccaac aacaaccgc ccaacgcacac cgagaccctc 1200
 cgcggccgcg gcccgaacat gaaggacaac tggcgcagcg agctgtacaa gtacaagggtg 1260
 gtgcgcatcg agccccctgg cgtggccccc acccaggcga acgcggcggt ggtgcagcgc 1320
 gagaagagcg cccgtgggcct ggggcctcg ttcatcggt tcctgggcgc cgccgggagc 1380
 accatggcgcc cccgcctccgt gaccctgacc gtgcaggccc gccagctgtc gageggcatac 1440
 gtgcagcagc agaacaacct gctgcgcgcc atcgaggccc agcagcactc gctgcagctg 1500
 accgtgtggg gcatcaagca gctgcaggcc cgcacatctgg cctggagcg ctacctgaag 1560
 gaccaggcgc tgctgggcat ctgggctgc agcggcaagc tgatctgcac caccaccgtg 1620
 ccctggaaaca gcagctggag caacaagagc ctgaccgaga tctgggacaa catgacctgg 1680
 atggagtggg agcgcgagat cggcaactac accggcctga tctacaacct gatcgagatc 1740
 gcccagaacc agcaggagaa gaacgagcag gagctgtgg agctggacaa gtggccagc 1800
 ctgtggaaact ggttcgacat caccaactgg ctgtggtaca tctaagatat cgatccctct 1860
 aga
 1863

<210> 64

<211> 2634

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: gp160.modUS4

<400> 64

gaattcgcca ccatggatgc aatgaagaga gggtctgtct gtgtgtgtgt gctgtgtgg 60
 gcagtttctcg ttcgccccag cgccaccacc gtgtgtgg tgaccgtgtta ctacggcggt 120
 cccgtgtgga aggaggccac caccacccctg ttctgcgcac ggcacgcacaa ggcttacaag 180
 gcccggggcc acaacgtgtg ggccacccac gcgtgcgtgc ccaccgaccc caaccccccag 240
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catgggtggag 300
 cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
 accccccctgt gcgtgaccct gaactgcacc gacaagctga cccggcagcac caacggcacc 420
 aacagcacca gcccacccaa cagcaccacgc ggcaccaaca gcaccagcac caacagcacc 480
 gacagctggg agaagatgcc cgagggcgag atcaagaact gcagcttcaa catcaccacc 540
 agcgtgcgcg acaagggtgca gaaggaggtac agcctgttct acaagctgca cgtgggtggcc 600
 atcgacaacg acaacgcaccc ctacccgcctg atcaactgca acaccagcgat gatcaccacc 660
 gcctggccca agtgtgagctt cgagcccatc cccatccact actgcgcgcc cggccggcttc 720
 gccatccctga agtgcacccaa caagaagttt aacggcaccgc gcccctgcac gacgtgagc 780
 accgtgcagt gcacccacgg catccgcggcc gtggtgagca cccagctgt gctgaacggc 840
 agcctggccg aggaggagat cgtgcgcgc tccgagaact tcaccgacaa cggccaaagacc 900
 atcatcggtc agtgcacccaa gtcgggtggat atcaactgca tccggcccaaa caacaacacg 960
 cgtaaaggcgc tccacatccgg cccggccgc gccttctaccc ccaccggcga catcatcg 1020
 gacatccgcg agccccactg caacatcgac aaggccaaact ggaccaacac cctcgagcag 1080
 atcgatggaga agtgcgcgc gacgttcggc aacaacaaga ccatcatctt caacagcagc 1140
 agcggccggcg acccccgagat cgtttccac agtcaact gccggccggca gttcttctac 1200
 tgcacccacca gccagctgtt caacagcacc tggaccaatca cccaggagggt gacaaagacc 1260
 aaggagaacg acaccatcat cctgcctgc cgcacatccgcgca agatcatcaa catgtggcag 1320

ggagggtgggca	aggccatgt	cgcccccccc	atccgcggcc	agatcaagt	cagcagcaat	1380
attaccggcc	tgcgtctgac	ccgcgcacggc	ggcaccaaca	acaaccgcac	caacgcacacc	1440
gagaccttcc	ccccggcg	cggcaacatg	aaggacaact	ggcgca	gctgtacaag	1500
tacaagggtgg	tgcgtatcg	gccccctggc	gtggccccca	cccaggccaa	gcccgcgtg	1560
gtgcagcgcg	agaagcgcgc	cgtgggcctg	ggcgcctgt	tcatcggtt	cctggcgcc	1620
gccgggagca	ccatgggcgc	cgcctccgt	accctgaccg	tgcaggcccc	ccagctgt	1680
agcggcatcg	tgcagcagca	gaacaacctg	ctgcgcgc	tgcaggcccc	gcagcacctg	1740
ctgcagctga	ccgtgtgggg	catcaagcag	ctgcaggcccc	gcatcctggc	cgtggagcgc	1800
tacctgaagg	accagcagct	gctgggcata	tggggctgca	gcccgaagct	gatctgcacc	1860
accaccgtgc	ccttggAACAG	cagctggagc	aacaagagcc	tgaccgagat	ctgggacaac	1920
atgaccttgg	tggagtggg	gcccgcagatc	ggcaactaca	ccggcctgat	ctacaacctg	1980
atcgagatcg	cccagaacca	gcaggagaag	aacgagcagg	agctgtgg	gctggacaag	2040
tggggccagcc	tgttggAACTG	gttcgcacatc	accaactggc	tgtggtacat	ccgcacatttc	2100
atcatgatcg	tggggccgt	gatcgccctg	cgcacatgtgt	tgcgcgtgt	gagcatcg	2160
aaccgcgtgc	gccagggtta	cagccccatc	agcctgcaga	ccggcctg	cggccagcgc	2220
ggcccccgacc	gccccggagg	catcgaggag	gagggcggcg	agcgcgaccg	cgaccgcagc	2280
aaccgcctgg	tgcacggcct	gctggccctg	atctggacg	acctgcgcag	cctgtgcctg	2340
ttcagctacc	accgcctgcg	cgacctgt	ctgatcg	tggccatcg	ggagctgt	2400
ggccgcgcg	gcttggaggg	cctgaagtac	tggtggaaacc	tgctgcagta	ctggaggccag	2460
gagctgaaga	gcagcgcgt	gagcctgt	aacgcacccg	ccatcgccgt	ggccgagggc	2520
accgaccgca	tcatcgagat	cgtgcagcgc	atcttccgcg	ccgtgtatcca	catccccgc	2580
cgcatccgc	aggccctgg	gcccgcctg	ctgttaagata	tcggatcctc	taga	2634

<210> 65

<211> 2538

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

Description of file

<400> 65

gaatccggca ccatggatgc aatgaagaga gggctctgct gtgtctgtct gctgtgtgga 60
gcagtctcg ttcgcggccag cgccaccacc gtgtctgtgg tgaccgtgtta ctacggcgtg 120
cccgtgtgga aggaggccac caccaccctg ttctgcggca gcgacgcca ggcttacaag 180
ggcgaggccc aacaacgtgtg ggcacccac gcctgcgtgc ccaccgaccc caaccccccag 240
gagggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgag 300
cagatgcatttggagacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gaactgcacc gacaagctgg ggcggccgg cgagatcaag 420
aactgcagct tcaacatcac caccagcgtg cgcgacaagg tgagaaggaa gtacagctg 480
ttctacaacgc tggacgtggc gccccatcgac aacgacaacg ccagctaccg cctgatcaac 540
tgcaacacca gcgtgtatcac ccaggcctgc cccaaaggta gcttcgagcc catccccatc 600
cactactgcg ccccccggg cttcgccatc ctgaagtgcaggacaagaa gttcaacggc 660
accggccctt gcaagaacgt gaggcaccgtg cagtgcaccc acggcatccg ccccggtgtg 720
agcacccacgc tgctgtgaa cggcagcctg gccgaggagg agatctgtct ggcgtccgag 780
aacttcacccg acaacgcca gaccatcatc gtgcagctga acgagtcctg ggagatcaac 840
tgcatccggcc ccaacaacaa cacgcgtaaag agcatccaca tcggcccccgg cccgcgccttc 900
tacgcccacccg gcgcacatcat cggcgacatc cggcaggcccc actgcaacat cagcaaggcc 960
aactggacca acaccctcga gcagatctgt gagaagctgc gcgagcagg tggcaacaac 1020
aagaccatca tcttcaacag cagcagcggc ggcgaccccg agatctgttt ccacagcttc 1080
aactgcggcg gcgagttctt ctactgcac accagccagc tggtaacacag caccctggaaac 1140
atcaccggagg aggtgaacaa gaccaaggag aacgacacca tcatcctgcc ctggcgatcc 1200
cgccagatca tcaacatgtg gcaggaggtt ggcaaggcca tgtacgcccc ccccatccgc 1260
ggccagatca agtgcagcag caatattacc ggcctgtgc tgaccccgca cggcggcacc 1320
aacaacaacc gcaccaacga caccggagacc ttccggcccg gggcggcaaa catgaaggac 1380
aactggcgca gcgagctgtt caagtacaag gtggtgccca tegagccccctt gggcgtggcc 1440
cccaccccaagg ccaagcggcg cgtggtgcaag cgcgagaacg ggcggctggg cctggcgcc 1500
ctgttcatcg gtttcttggg cggccggggg agcaccatgg ggcggcctc cgtgaccctg 1560

```

accgtgcagg cccgccagct gctgagcggc atcgtgcagc agcagaacaa cctgctgcgc 1620
gccatcgagg cccagcagca cctgctgcag ctgaccgtgt ggggcatcaa gcagctgcag 1680
gccccgcatcc tggccgtgga gcgctacctg aaggaccagc agctgctggg catctggggc 1740
tgcagcgca agctgatctg caccaccacc gtgcctgga acagcagctg gagcaacaag 1800
agcctgaccg agatctggga caacatgacc tggatggagt gggagcgcga gatcggcaac 1860
tacaccgccc tgatctacaa cctgatcgag atcgcccaga accagcagga gaagaacgag 1920
caggagctgc tggagctgga caagtggcc agcctgtgga actggttcga catcaccaac 1980
tggctgttgt acatccgcat cttcatcatg atcgtggcg gcctgatecg cctgcgcata 2040
gtgttcggcg tgctgagcat cgtgaaccgc gtgcggcagg gtacagccc catcaggctg 2100
cagaccgcgc tgeccgcccc gcgcggcccc gaccggcccc agggcatcga ggaggagggc 2160
ggcgagcgcg acccgacccg cagcaaccgc ctggtgcacg gcctgtggc cctgatctgg 2220
gacgacctgc gcagcctgtg cctgttcage taccaccgc tgcgacact gctgctgatc 2280
gtggccgcga tcgtggagct gctggccgc cgcggtctggg aggccctgaa gtactggttg 2340
aacctgctgc agtactggag ccaggagctg aagagcagcg ccgtgagcct gttcaacgcc 2400
accgccccatcg ccgtggccga gggcaccgac cgcatcatcg agatcgtgca ggcacatttc 2460
cgcgccgtga tccacatcccc cccgcgcata cgccaggccc tggagcgcgc cctgctgtaa 2520
gatatcgatc cctctaga

```

<210> 66
<211> 2553

<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp160.modUS4.delV2

<400> 66

gaattcgcca	ccatggatgc	aatgaagaga	gggctctgct	gtgtgctgct	gctgtgtgga	60
gcagtcttcg	tttcgcccag	cgccaccacc	gtgctgtggg	tgaccgtgt	ctacggcgtg	120
cccgtgtgga	aggaggccac	caccacccctg	ttctcgccca	gcgacgcca	ggcttacaag	180
gccgaggccc	acaacgtgt	ggccaccacc	gctcgctgc	ccaccgaccc	caacccccag	240
gaggtgaacc	tgaccaacgt	gaccgagaac	ttcaacatgt	ggaagaacaa	catgggtggag	300
cagatgcatt	aggacatcat	cagcctgtgg	gaccagagcc	tgaagccctg	cgtgaagctg	360
accccccgt	gcgtgaccct	gaactgcacc	gacaagctga	ccggcagcac	caacggcacc	420
aacagcacca	gcccaccaa	cagcaccaggc	ggcaccaaca	gcaccagcac	caacagcacc	480
gacagctggg	agaagatgcc	cgagggcgag	atcaagaact	gcagcttcaa	catcgccgccc	540
ggccgcctga	tcaactgcaa	caccagcgta	atcaccctagg	cctgccccaa	ggtgagttc	600
gagcccatcc	ccatccacta	ctgcgcccc	gccggcttcg	ccatcctgaa	gtgcaaggac	660
aagaagtta	acggcacccgg	ccccctgcaag	aacgtgagca	ccgtgcagtg	cacccacggc	720
atccgcctcg	tggtgagcac	ccagctgtcg	ctgaacggca	gcctggccga	ggaggagatc	780
gtgctgcct	ccgagaactt	caccgacaac	gccaagacca	tcatcgtca	gctgaacagg	840
tccgtggaga	tcaactgcat	ccggccccaac	aacaacacgc	gtaagagcat	ccacatcgcc	900
ccggcccgcg	ccttctacgc	caccggcgac	atcateggcg	acatccgcca	ggccactgtc	960
aacatcagca	aggccaactg	gaccaacacc	ctcgagcaga	tcgtggagaa	gctgcgcgag	1020
cagttcggca	acaacaagac	catcatcttc	aacagcagca	gcggcggcga	ccccgagatc	1080
gtgttccaca	gcttcaactg	cgggcccgag	ttcttctact	gcaacaccag	ccagctgttc	1140
aacagcacct	ggaacatcac	cgaggaggtg	aacaagacca	aggagaacga	caccatcata	1200
ctggccctgcc	gcatccgcca	gatcatcaac	atgtggcagg	aggtgggcaa	ggccatgtac	1260
gcccccccca	tcccgccgca	gatcaagtgc	agcagcaata	ttacccggct	gctgctgacc	1320
cgcgacggcg	gcaccaacaa	caaccgcacc	aacgcacccg	agaccttccg	ccccggccgc	1380
ggcaacatga	aggacaactg	gchgagcgag	ctgtacaagt	acaagggtgt	ggcgcattcgag	1440
ccccctggcg	tggcccccac	ccaggccaa	cgccgcgtgg	tgacgcgcga	gaagcgcc	1500
gtgggcctgg	gcccctgtt	catcggttc	ctggggcccg	ccggggagcac	catgggcgccc	1560
gcctccgtga	ccctgaccgt	gcaggcccg	cagctgtga	gccccatcgt	gcagcagcag	1620
aacaacctgc	tgcgcgccat	cgaggcccg	cagcacctgc	tgacgctgac	cgtgtgggc	1680
atcaaggcgc	tgcaggcccc	catccggcc	gtggagcgct	acctgaagga	ccagcagctg	1740
ctgggcattt	ggggctgca	cgcaagctg	atctgcacca	ccaccgtgcc	ctggaaacagc	1800
agctggagca	acaagagcct	gaccgagatc	tgggacaaca	tgacctggat	ggagtggqaq	1860

cgcgagatcg gcaactacac cggcctgatc tacaacctga tcgagatcgc ccagaaccag 1920
 caggagaaga acgagcagga gctgctggag ctggacaagt gggccagcct gtggaaactgg 1980
 ttcgacatca ccaactggct gtggatcac tcatgtatcg 2040
 atcggcctgc gcatacggtt cggcgtctg agcatcgta accgcgtgcg ccagggctac 2100
 agccccatca gcctgcagac cggcctgccc gcccagecg 2160
 atcgaggagg agggcggcga gcgcgaccgc gaccgcagca accgcctggt gcacggcctg 2220
 ctggccctga tctgggacga cctgcgcagc ctgtgcctgt tcagettacca cccgcctgcgc 2280
 gacctgctgc tgatcgatggc cccatcgatg gagctgtgg gcccggcgg ctgggaggcc 2340
 ctgaagtact ggtggaaact gctgcgtac tgagggcagg agtgaagag cagcgccgtg 2400
 agcctgttca acgcccaccc catcgccgtg gccgaggggca cccaccat catcgagatc 2460
 gtgcagcgca tttcccgcc cgtatccac atccccccgc gcatccgcca gggcctggag 2520
 cgcgcctgc tgtaagatat cggatcctct aga 2553

<210> 67

<211> 2340

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp160.modUS4.delV1/V2

<400> 67

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgg 60
 gcagtcttcg ttccggccag cggcaccacc gtgctgtgg tgaccgtgta ctacggcgtg 120
 cccgtgtgga aggaggccac caccaccctg ttctgcgcca ggcacgcggaa ggcttacaag 180
 gcccggggcc acaacgtgtg ggcacccac gcctgcgtgc ccaccgcacc caaccccccag 240
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
 cagatgcgtg aggacatcat cggcctgtgg gaccagagcc tgaaggcctg cgtggggcgcc 360
 ggcaggcct gccccaaagggt gagcttcgag cccatccccca tccactactg cggcccccggcc 420
 ggcttcgcca ttctgaagtg caaggacaag aagttcaacg gcacccggccc ctgcaagaac 480
 gtgagcaccg tgcgtgcac ccacggcattc cggccctgtgg tgagcaccca gctgctgtg 540
 aacggcagcc tggccgagga ggagatcgatg ctgcgtcccg agaacttccac cgacaacgc 600
 aagaccatca tcgtgcagct gaacgatcc tgagggatca actgcattccg ccccaacaac 660
 aacacgcgtt agagcatcca catcgcccc gggcgcgcct tctacgcac cggcgcacatc 720
 atcggcgaca tccgcccaggc ccactgcaac atcagcaagg ccaactggac caacaccctc 780
 gacgagatcg tggagaaggt ggcgcgacgat ttcggcaaca acaagaccat catettcaac 840
 agcagcagcg gcccgcaccc cgagatcgatg ttccacagct tcaactgcgg cggcggatcc 900
 ttctactgca acaccagccaa gctgttcaac agacacccatggaa acatcaccga ggaggtgaac 960
 aagaccaagg agaacgcacac catcatccgt cccgcgcgc tccgcccagat catcaacatg 1020
 tggcaggagg tggcaaggc catgtacgatcc ccccccattcc gcccggcagat caagtgcac 1080
 agcaatatta cccgcctgt gctgacccgc gacggcggca ccaacaacaa cccgcaccaac 1140
 gacccgaga cttccgcggc cggcggcggc aacatgaagg acaactggcg cagcggatcg 1200
 tacaagtata aggtgggtcg catcgatccctg ctggggcgtgg ccccccacca gggcaagcgc 1260
 cgcgtgtgc agcgcgagaa ggcgcgcgtg ggcctggcgc ccctgttcat cggcttcctg 1320
 ggcgcgcgcgg ggagccatcat gggcgcgcgc tccgtgaccc tgaccgtgca gggccgcag 1380
 ctgctgagcg gcatacgatgc gcagcagaac aacatcgatgc ggcgcattcgaa gggcccgac 1440
 cacctgtgc agctgaccgt gtggggcattc aacatcgatgc aggccgcattc cctggccgtg 1500
 gagcgctacc tgaaggacca gacatcgatg ggcattcgatgg gctgcggcgg caagctgtatc 1560
 tgcacccatca cctgtccctg gaacacgcgc tggagcaaca agacccatcgac cggatctgg 1620
 gacaacatga cttggatggaa gtggggcgcg gagatcgatc actacacccg cctgtatctac 1680
 aacctgtatcg agatcgccccca gaaccgcgc gagaagaacg agcggggatc gctggggatcg 1740
 gacaagtggg ccagcctgtg gaactggatcc gacatccatca actggctgtg gtacatccgc 1800
 atcttcatca tgatcgatggg cggcctgtatc ggcctgcgc tctgtttcgat cgtgtgcg 1860
 atcgtgaacc gcgtgcgcctt gggctacatcg cccatcgatc tgcagacccg cctggccgc 1920
 cagcgcggcc cccatcgatc cggcgcgcgtg gggccatcgatc gggacccatcgatc 1980
 cgcgcgcgc ggcgcgcgc ggcgcgcgc ggcgcgcgc ggcgcgcgc ggcgcgcgc 2040
 tgcctgttca gtcaccaccg cctgcggcgc tgcgtgcgtatc tgcgtggcccg catcgatggag 2100
 ctgctggcc ggcgcggctg ggaggccctg aagtactggatc ggaactcgatc cgtactgg 2160

agccaggagc tgaagagca gccccgtgagc ctgttcaacg ccaccgcatt cggcggtggcc 2220
 gagggcaccg accgcattcat cgagatcgta cagcgcatct tccgcgcgt gatccacatc 2280
 ccccgccgca tccgcccaggc cctggagcgc gcccgtgt aagatatacg atccctctaga 2340

<210> 68
 <211> 2385
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 gp160.modUS4del 128-194

<400> 68
 gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgtgtgt gctgtgtgga 60
 gcagtcgttgc ttgcgtcccg cgcacccacc gtgtgtggg tgaccgtgtta ctacggcggtg 120
 cccgtgtggaa aggaggccac caccacccgt ttctgcgtcc ggcacgcacaa ggcttacaag 180
 gcccaggcccc acaacgtgtg ggcacccac gcgtgcgtgc ccacccgaccc caaccccccag 240
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtggag 300
 cagatgcatttgg aggcacatcat cagcgtgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
 accccctgtt gctgtggggc agggaaactgc gagaccagcg tgatcaccca ggcctgcccc 420
 aagggtgagct tcgagcccat ccccatccac tactgcgtcc cgcgggtt cgcacatccgt 480
 aagtgcagg acaagaagtt caacggcacc ggccctgtca agaacgtgag caccgtgcag 540
 tgcacccacg gcatccgccc cgtgggtgagc acccagctgc tgctgaacgg cagcctggcc 600
 gaggaggaga tctgtgtgtgg ctccgagaac ttcaaccgaca acgccaagac catcatcg 660
 cagctgaacg agtccgtgga gatcaactgc atccggccca acaacaacac gctgtgggg 720
 atccacatcg gcccggccg cgccttctac gccaccggcg acatcatcg cgacatccgc 780
 caggcccact gcaacatcg caaggccaaac tggaccaaca ccctcgagca gatcgtggag 840
 aagctgcgcg agcagttcg cacaacaag accatcatct tcaacagcag cagcggccgc 900
 gaccccgaga tctgttccaa cagcttcaac tgccggccgc agttttctca ctgcaacacc 960
 agccagctgt tcaacagcacc ctggaacatc accgaggagg tgaacaagac caaggagaac 1020
 gacaccatca tctgtccctg cgcacccgc cagatcatca acatgtggca ggaggtggc 1080
 aaggccatgt acggccccc catccgcgc cagatcaagt gcagcagcaa tattaccggc 1140
 ctgtgtgtga cccgcgacgg cggcacaac aacaaccgc ccaacgacac cgagacccttc 1200
 cgcggccgc gcccggccat gaaggacaac tggcgcagcg agctgtacaa gtacaagggt 1260
 gtgcgcacatcg agcccttggg cgtggccccc acccaggccca acgcggcggt ggtcagcgc 1320
 gagaagcgcg cctgtggccct gggccctgt ttcatcggtc tccctggccgc cgccgggagc 1380
 accatggcgccgc cccctccgt gaccctgtacc gtgcaggccc gccagctgt gacggccatc 1440
 gtgcagcagc agaacaacct gctgcgcgc atcgaggccc acgcggccatc gtcagctgt 1500
 accgtgtggg gcatcaagca gctgcaggcc cgcacccctgg ccgtggagcg ctacctgaag 1560
 gaccagcagc tgctggccat ctggggctgc agcggcaagc tgatctgcac caccaccgt 1620
 ccctggaaaca gcaagctggag caacaagagc ctgaccgaga tctgggacaa catgaccctgg 1680
 atggagtggg agcgcgagat cggcaactac accggcctga tctacaacct gatcgagatc 1740
 gcccagaacc agcaggagaa gaacgagcag gagctgtgtgg agctggacaa gtggggccagc 1800
 ctgttgaact gtttcgacat caccaacttgg ctgtgttaca tccgcacatc catcatgatc 1860
 gtggggccgc tgatcggtt ggcacatcg ttcgcgtgc tgagcatgt gacccgtg 1920
 cggccaggccat acagcccatc cggcctgcgacc accccgcgtc cggcccgccgc cggcccccgc 1980
 cgcggccgc gcatcgaggaa ggaggccgc gacccggccgc gacccggccgc caaccgcctg 2040
 gtgcacggcc tctgtggccct gatctgggac gacccgtgc gacccgtgc gctgtgtcc gttcagctac 2100
 caccgcctgc gacccgtgt gtcgtgtcc gccgcacccgt tggagctgt gggccggccgc 2160
 ggctggggagg ccttgaagta ctggtggaaac ctgtgcgtactggagcca ggagctgaag 2220
 agcagcgcgc tgagctgtt caacgcacc gccatcgccg tggccggagg caccgaccgc 2280
 atcatcgaga tcgtgcagcg catccgcgc gccgtgtatcc acatcccccgc cccgcacccgc 2340
 cagggccctgg agcgcgcctt gctgtaaat atccgttctt cttaga 2385

<210> 69
 <211> 144
 <212> DNA
 <213> Human immunodeficiency virus

<400> 69
gacaccatca tcctgcctcg ccgcattccgc cagatcatca acatgtggca ggaggtggc 60
aaggccatgt acgccccccc catccgcggc cagatcaagt gcagcagcaa catcacccgc 120
ctgctgctga cccgcgacgg cgcc 144

<210> 70
<211> 144
<212> DNA
<213> Human immunodeficiency virus

<400> 70
ggaactatca cactcccatg cagaataaaa caaattataa acaggtggca ggaagtagga 60
aaagcaatgt atgccccctcc catcagagga caaatttagat gctcatcaa tattacagga 120
ctgcttattaa caagagatgg tggt 144

<210> 71
<211> 144
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic Env
US4 common region

<400> 71
gacaccatca tcctgcctcg ccgcattccgc cagatcatca acatgtggca ggaggtggc 60
aaggccatgt acgccccccc catccgcggc cagatcaagt gcagcagcaa catcacccgc 120
ctgctgctga cccgcgacgg cgcc 144

<210> 72
<211> 144
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic Env
SF162 common region

<400> 72
ggcaccatca ccctgcctcg ccgcattcaag cagatcatca accgctggca ggaggtggc 60
aaggccatgt acgccccccc catccgcggc cagatccgt gcagcagcaa catcacccgc 120
ctgctgctga cccgcgacgg cgcc 144

<210> 73
<211> 4766
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
gp160.modUS4.gag.modSF2

<400> 73
gaattcgcca ccatggatgc aatgaagaga gggctctgt gtgtgtgtct gctgtgtgga 60
gcagtcttcg tttcgccag cgccaccacc gtgtgtgg tgaccgtgtta ctacggcgtg 120
cccggtgga aggaggccac caccaccctg ttctgcggca gcgcgcggaa ggcttacaag 180
gccgaggccc acaacgtgtg ggccaccac gcctgcgtgc ccaccgcacc caaccccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgag 300
cagatgcattt aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360

accccctgt gggtgaccct gaactgcacc gacaagctga cccggcagcac caacggcacc 420
 aacagcacca gcggcaccaa cagcaccaggc ggcaccaaca gcaccagcac caacagcacc 480
 gacagctggg agaagatgcc cgagggcgag atcaagaact gcagcttcaa catcaccacc 540
 agcgtgcgcg acaagggtca gaaggagtac agctgttct acaagctgga cgtggtgc 600
 atcgacaacg acaacgccc ctaccgctg atcaactgca acaccaggcgt gatcacccag 660
 gcctgccccca aggtgagctt cgagcccatc cccatccact actgcgc 720
 gccatctga agtgcagga caagaagtcc aacggcacccg gcccctgcaaa gaacgtgagc 780
 accgtgcagt gcacccacgg catccggcccc gtggtagca cccagctgt gctgaacggc 840
 agcctggccg aggaggagat cgtgctgcgc tccgagaact tcaccgacaa cgccaagacc 900
 atcatcgtc agctgaacga gtccgtggag atcaactgca tccggcccaa caacaacacg 960
 cgtaagagca tccacatcg ccccgccccgc gccttctacg ccacggcga catcatcgcc 1020
 gacatccgccc aggcccactg caacatcagc aaggccaaact ggaccaacac cctcgagcag 1080
 atcgtggaga agctgcgcga gcagttcgcc aacaacaaga ccatcatttt caacagcagc 1140
 agcggccgcg acccccggat cgtttccac agettcaact gcccggcga gttttctac 1200
 tgcaacacca gccagctgtt caacagcacc tggaacatca cccaggaggtt gaacaagacc 1260
 aaggagaacg acaccatcat cctgcccgtc cgcatccggc agatcatcaa catgtggcag 1320
 gaggtggca aggccatgtc cggccccccc atccggggcc agatcaagt cagcagcaat 1380
 attacggcc tgctgctgac cccgcacggc ggcaccaaca acaaccgcac caacgacacc 1440
 gagaccttcc gccccggcg cggcaacatg aaggacaact ggcgcagcga gctgtacaag 1500
 tacaagggtgg tgcgcatgtc gcccctggc gtggccccca cccaggccaa ggcgcgcgtg 1560
 gtgcagcgcg agaagcgcgc cgtggccctg ggcggccctgt tcatcggtt cctggcgcc 1620
 gccggggaca ccatggggcgc cgcctccgtg accctgaccg tgcaggcccc ccagctgctg 1680
 agcggcatcg tgcagcagca gaacaacctg ctgcgcgcga tcgaggcccc gcagcacctg 1740
 ctgcagctga cccgtgtgggg catcaagcagc ctgcaggcccc gcacccctggc cgtggagcgc 1800
 tacctgaagg accagcagct gctggccatc tgggctgca gcccggccatc gatctgcacc 1860
 accaccgtgc cctggaaacag cagctggagc aacaagagcc tgaccgagat ctggacaac 1920
 atgacctgga tggagtggga ggcgcagatc ggcaactaca cccgcctgtat ctacaacctg 1980
 atcgagatcg cccagaacca gcaggagaag aacgagcagg agctgctgga gctggacaag 2040
 tggccagcc tgggactatc accaactggc tgggtacat ccgcattttt 2100
 atcatgatcg tggggggct gateggccctg cgcacccgtt tgcaggccctg gaggatcg 2160
 aaccgcgtgc gccagggtta cagccccatc agcctgcaga cccgcctgca cggccagcgc 2220
 gggccggacc gccccggagg catcgaggag gaggccggcg agcgcgaccg cggccggacc 2280
 aaccgcctgg tgcacggcc tgcacggccatc gtcggccctg atctgggacg acctgcgcag cctgtgcctg 2340
 ttcaagttacc accgcctgcgc cgcacccgtt ctgatcgatc cccgcacccgtt ggagctgctg 2400
 ggcggccgcg gtcggggaggc cctgaagtac tgggtggacc tgctgcagta ctggagccag 2460
 gagctgaaga gcaagcgcgt gaggccgttca aacgcacccg ccacccgtt ggcggaggcc 2520
 accgaccgca tcatcgagat cgtgcagcgc atctccgcg ccgtgatcca catccccccgc 2580
 cgcacccgcg aggccctggc ggcgccttgc ctgtaaatata tcggatccctc tagagaattc 2640
 cggccccccc cccccccccc ctctccctcc ccccccccta acgttactgg cccaaaggccgc 2700
 ttggataataag gccgggtgtgc gtttgttat atgttattttt ccaccatatt ggcgtctttt 2760
 ggcaatgtga gggccggaa acctggccct gtcttctgtc cgagcatcc taggggtctt 2820
 tccccctctg cccaaaggaaat gcaaggctgt ttgaatgtcg tgaagggaaatc agttccctctg 2880
 gaagcttctt gaagacaaaac aacgtctgtc ggcacccctt gcaggccagcg gaacccccc 2940
 cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc tgcaaaaggcg 3000
 gcacaacccc agtgccacgt tggatgtgg atagttgtgg aaagagtcaa atggctctcc 3060
 tcaagcgat tcaacaaggaa gctgaaggat gcccggccatc taccatcgatc tatggatct 3120
 gatctggggc ctcgggtgcac atgctttaca tgggtttatcg cggatgttaaa aaaacgtcta 3180
 gggccccccgc accacggggc cgtgggttttcc tttggaaaaa cacgataata ccatggggcgc 3240
 cccgcggccaggc gtcgtgcgcg gggccggaggat ggacaatgtgg gagaagatcc gcctgcgc 3300
 cggccggcaag aagaagtaca agctgaagca catcgatcgatc gcccggcccg agctggagcg 3360
 cttcgccgtg aaccccccggcc tgcggggagc cagcgaggcc tgcggccaga tcctggggcca 3420
 gtcgcagccccc agcctgcaga cccgcacccgcg ggacgtgcgc acaccgtggc 3480
 caccctgtac tgcgtgcacc agcgcacccgc cgtcaaggac accaaggagg ccctggagaa 3540
 gatcgaggag gaggcagaaca agtccaaagaa gaaggccccag caggccggcccg cccggccgg 3600
 caccggcaac agcagccagg tgagccagaa ctacccatc tgccagaacc tgccaggcc 3660
 gatggtgac caggccatca gccccccgcac cctgaacccgc tgggtgaagg tgggtggagg 3720
 gaaggcccttc agccccgggg tggatccccat gttcaggccgc ctgagcggagg ggcacccccc 3780
 ccaggacccgt aacacgtat tgaacaccgtt gggccggccac caggccggccca tgccatgtct 3840
 gaaggagacc atcaacggagg aggccggccgc gtggggaccgc tgccacccccc 3900

ccccatcgcc cccggccaga tgcgcgagcc ccgcggcagc gacatcgccg gcaccaccag 3960
 caccctgcag gaggcagatcg gctggatgac caacaacccc cccatccccg tgggcgagat 4020
 ctacaagcgg tggatcatcc tgggcctgaa caagatcgta cgatgtaca gccccaccag 4080
 catcctggac atccgcagg gccccaaagga gccctccgc gactacgtgg accgcttcta 4140
 caagaccctg cgcgctgagc aggccagcca ggacgtgaag aactggatga ccgagaccct 4200
 gctggtcag aacgccaacc cgcactgca gaccatctg aaggctctcg gccccggc 4260
 caccctggag gagatgtga cgcctgcca gggcgtggc ggccccggcc acaaggcccc 4320
 cgtgtggcc gaggcgtga gccaggtgac gaaccggcg accatcatga tgcagcgcgg 4380
 caactccgc aaccagcgga agaccgtcaa gtgttcaac tgccgcaagg agggccacac 4440
 cggcaggAAC tgcgcgccc cccgcaagaa ggctgtctgg cgctgcggcc gcgagggcca 4500
 ccagatgaag gactgcaccc agcgccaggc caacttctg ggcaagatct gccccagacta 4560
 caaggccgc cccggcaact tcctgcagag cgcctggcag cccaccggcc ccccccggagga 4620
 gagttccgc ttccggcgagg agaagaccac ccccgccag aagcaggagc ccatcgacaa 4680
 ggagctgtac cccctgacca gcctgcgcag cctgttcggc aacgaccctca gcagccagta 4740
 agaattcaga ctcgagcaag tctaga 4766

<210> 74

<211> 4689

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
gp160.modsF162.gag.modsF2

<400> 74

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
 gcagtcttcg ttccggccag cgcgtggag aagctgtgg tgaccgtgt a taccggcgtg 120
 cccgtgtgga aggaggccac caccacccctg ttctgcgcca ggcacgc当地 ggcctacgac 180
 acccgagggtgc acaacgtgtg ggcacccac gcgtgcgtc ccaccgaccc caacccccc 240
 gagatcgatgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catgtggag 300
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaaggccctg cgtgaagctg 360
 accccctgt gctgtgaccct gcaactgc当地 aacctgaaga acgccc当地 caccacgac 420
 agcaactgga aggagatgga cgcggcgag atcaagaact gcagcttcaa ggtgaccacc 480
 agcatccgca acaagatgca gaaggagtac gccc当地ttct acaagctgga cgtggtgc当地 540
 atcgacaacg acaacaccag ctacaagctg atcaactgca acaccagcgt gatcaccacc 600
 gcctgccccca aggtgagctt cgagccatc cccatccact actgc当地 cgc当地ggcttc 660
 gccatctga agtgc当地acgca caagaagttc aacggc当地eg gccc当地tc当地 acacgtgagc 720
 accgtgc当地gt gcacccacgg catccgcccc gttgtgagca cccagctgt gctgaacggc 780
 agcctggccg aggaggccgt ggtgatccgc agcgagaact tcaccgacaa cgcc当地agacc 840
 atcatcgatgc agctgaagga gacgtggag atcaactgca cccgccccaa caacaacacc 900
 cgcaagagca tcaccatcg cccggccgc gccttctacg ccaccggc当地 catcatcgcc 960
 gacatccgccc aggccc当地tgc当地 caacatcgac ggc当地gagaagt ggaacaacac cctgaagcag 1020
 atcgatgc当地ca agctgc当地ggc ccagttccgc aacaagacca tctgttcaa gcagagcagc 1080
 ggc当地ggccacc cccgagatcgat gatgc当地acg ttc当地actgca gccc当地gagtt cttctactgc 1140
 aacagcaccc agctgttcaa cagcaccctgg aacaacacca tggccccc当地 caacaccaac 1200
 ggc当地ccatca ccctgccccctg cgc当地atcaag cagatcatca accgctgca ggaggtggc 1260
 aaggccatgt acgcccccc catccgcccc cagatccgct gcagc当地acca catcaccggc 1320
 ctgtgtctga cccgcccccc cggc当地aggag atcagcaaca ccaccgagat cttccgcccc 1380
 ggc当地ggccg acatcgccgc caactgccc当地 agcgagctgt acaagttacaa ggtgtgag 1440
 atcgagcccc tgggc当地gtggc ccccaaccaag gcaagccgc gctgtgtgca ggc当地gagaag 1500
 cgc当地cgatgc ccctggccgc catgttctg ggttctctgg gccc当地ccgg cagcaccatg 1560
 ggc当地ccggca gcctgaccct gaccgtgc当地 gccc当地ccagc tctgtgagccg catcgatgc 1620
 cagc当地acca acctgtcgcc cgc当地atcgag gccc当地ccagc acctgtgca gctgaccgtg 1680
 tggggcatca agcagctgca ggc当地ccgctg ctggccgtgg agcgatcttca gaaggaccag 1740
 cagctgtgg gcatctgggg ctgc当地ccggc aagctgtatct gcaccacccgc cgtccctgg 1800
 aacgccc当地gt ggagcaacaa gacgtggac cagatctgg acaacatgac ctggatggag 1860
 tgggc当地ggccg agatcgacaa ctacaccaac ctgtatctaca ccctgatgc当地 ggagagccag 1920
 aaccagcagg agaagaacga gcaggagctg ctggagctgg acaagtgccg cagcctgtgg 1980

aactggttcg acatcagcaa gtggctgtgg tacatcaaga tcttcatcat gatcgtgggc 2040
 ggcctggtgg gcctgcgcac cgtgttccacc gtgctgagca tcgtgaaccg cgtgcgccag 2100
 ggctacagcc ccctgagctt ccagacccgc ttccccgccc cccgcggccc cgaccgcccc 2160
 gagggcatcg aggaggaggg cggcgagcgc gaccgcgacc gcagcagccc cctggtgac 2220
 ggcctgtgg ccctgatctg ggacgacctg cgcagcctgt gcctgttcag ctaccaccgc 2280
 ctgcgcacc tgatcctgtat cggcccccgc atcgtggagc tgctggggcc cgcggctgg 2340
 gagggcctga agtactgggg caacctgctg cagtaactggta tccaggagct gaagaacagc 2400
 gccgtgagcc tgttcgacgc catcgccatc gccgtggccg agggcaccca ccgcatacatc 2460
 gaggtggccc agcgcategg cgcgccttc ctgcacatcc cccgcgcac cgcgcaggc 2520
 ttcgagcgc ccctgctgtat actcgagcaa gtctagagaa ttccgcggcc cccccccccc 2580
 cccctctccc tccccccccc ctaacgttac tggccgaagc cgcttgaat aaggccggtg 2640
 tgcgtttgtc tatatgttat tttccaccat attggcgctt tttggcaatg tgagggcccg 2700
 gaaacctggc cctgtcttc tgacgagcat tcttaggggt cttcccttc tcgccaagg 2760
 aatgcaaggt ctgttgaatg tcgtgaagga agcagtccct ctggaaagctt cttgaagaca 2820
 aacaacgtct gttagcgcaccc ttgcaggca gccggaaaaaa ccacctggcg acaggtgcct 2880
 ctgcggccaa aagccacgtg tataagatac acctgcaaaag gggcacaac cccagtgcca 2940
 cgttgtgagt tggatagttt tggaaaaggt caaatggctc tcctcaagcg tattcaacaa 3000
 ggggctgaag gatgcccaga aggtacccca ttgtatggga tctgatctgg ggcctcggtg 3060
 cacatgctt acatgtgtt agtcgaggta aaaaaaaaaacgt ctaggcccc cgaaccacgg 3120
 ggacgtgggtt ttcccttggaa aaacacgata ataccatggg cggccgcgc agcgtgctga 3180
 gcccggcga gctggacaag tgggagaaga tccgcctgcg cccggccgc aagaagaagt 3240
 acaagctgaa gcacatcgat tggggccagcc gcgagctggat cgcgttcgcgttgcctgc 3300
 gcctgctggc gaccagcgcag ggctggccgc agatctggg ccagctgcag cccagctgc 3360
 agaccggcag cgaggagctg cgcagcctgt acaacaccgt ggccaccctg tactgcgtgc 3420
 accagcgcacat cgacgtcaag gacaccaagg aggccctggaa gaagatcgag gaggagcaga 3480
 acaagtccaa gaagaaggcc cagcaggccc cggccgcgc cggcaccggc aacagcagcc 3540
 aggtgagccaa gaactacccc atcgtgcaga acctgcagggtt ccagatgttgc caccaggcc 3600
 tcagccccccg caccctgaaac gcctgggtga aggtgggtggaa ggagaagcc ttcaagcccc 3660
 aggtgatccc catgttgcagc gcccctgagcg aggccgcaccc ccccccagac ctgaacacga 3720
 tggtaaacac cgtggggccgc caccaggccgc ccatgcagat gctgaaggag accatcaacg 3780
 aggaggccgc ctagtgggac cgcgtgcacc cctgtgcacgc cggcccccattt gccccccggcc 3840
 agatgcgcga gccccggcgc agegacatcg cggcaccac cagcaccctg caggagcaga 3900
 tcggctggat gaccaacaac ccccccattcc cctgtgggcga gatctacaag cggtggatca 3960
 tcctgggcctt gaacaagatc gtgcggatgt acagccccac cagcatctt gacatccgc 4020
 agggccccaag ggagcccttc cgcgactacg tggaccgctt ctacaagacc ctgcgcgtt 4080
 agcaggccag ccaggacgtt aagaactggta tgaccgagac cctgtgggtt cagaacgc 4140
 acccccactg caagaccatc ctgaaggctc tcggccccgc ggccaccctg gaggagatga 4200
 tgaccgcctg ccagggcgtt ggcggccccc gccacaaggc cggcgtgttgc gccgaggcga 4260
 tgagccaggat gacgaaccccg gcgaccatca ttagtgcagcg cggcaacttc cgcaaccagc 4320
 ggaagaccgtt caagtgtttc aactgcggca aggaggggcca caccggcagg aactgcgcgc 4380
 ccccccggcaa gaagggtgtt gggcgttgcg gcccgcgggg ccaccagatg aaggactgca 4440
 cccgagcgcacca ggcacaacttc ctggggcaaga tctggcccaatc ctacaaggcc cggccggca 4500
 acttcctgtca gagccggcccc gagcccacccg ccccccggca ggagagctt cgcgttgcgc 4560
 aggagaagac caccggccacca cagaagcagg agcccatcgaa caaggagctt taccggctga 4620
 ccagcctgttgc cagcctgttc ggcaacgacc ccagcagcc gtaagaattt agactcgagc 4680
 aagtctaga 4689

<210> 75

<211> 4472

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 gp160.modUS4.delV1/V2.gag.modSF2

<400> 75

gaatttcgcca ccatggatgc aatgaagaga gggctctgtt gtgtgtgtt gctgtgtgg 60
 gcagtcgttgc ttgcggccag cggccaccacc gtgtgtgtt gacccgttca ctacggcgtt 120

cccgtgtgga aggaggccac caccacctg ttctgcgcca gcgacgccaa ggcttacaag 180
 gccgaggccc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caacccccag 240
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgag 300
 cagatgcattt aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtggcgcc 360
 ggccaggcct gccccaaagggt gagcttcgg cccatccccca tccactactg cgcccccgcc 420
 ggcttcgcca tcctgaaggta caaggacaag aagttcaacg gcaccggccc ctgcaagaac 480
 gtgagcaccc tgcgtgcac ccacgcattt cgccccgtgg tgagcaccca gctgctgctg 540
 aacggcagcc tggccgagga ggagatctg ctgcgtccg agaacttcac cgacaacgc 600
 aagaccatca tcgtgcacgt gaacgatcc ttggagatca actgcattccg ccccaacaac 660
 aacacgcgtt agagcatcca catcgcccccc gggcgccct tctacgcac ccggcgacatc 720
 atcggcgaca tccgcaggcc ccactgcaac atcagcaagg ccaactggac caacaccctc 780
 gaggcagatcg tggagaaggt ggcgcagcc ttccggcaaca acaagaccat catcttcaac 840
 agcagcagcg gcccgcgaccc cgagatctg ttccacatgt tcaactgcgg cggcgagttc 900
 ttctactgca acaccagccaa gctgttcaac agacacccggaa acatcaccga ggaggtgaac 960
 aagaccaagg agaacgcacac catcatectg ccctgcccga tccggccat catcaacatg 1020
 tggcaggagg tggcgaaggcc catgtacggcc cccccccatcc gcccggat catgtgcagc 1080
 agcaatatta cccggctgtc gctgacccgc gacggcggca ccaacaacaa ccgcaccaac 1140
 gacaccgaga cttccgccc cggcgccggc aacatgaagg acaaactggc cagcgagctg 1200
 tacaagtaca aggtggtgcg catcgagccctt ctggggcgtgg cccccccacca ggccaagcgc 1260
 cgcgtgggtgc agcgcgagaa ggcgcggctg ggccctggggc ccctgttcat cgggttccctg 1320
 ggcgcggcccg ggagcaccat gggcgccggc tccgtgaccc tgaccgtgca gggccggccag 1380
 ctgctgagcg gcatcggtca gcaagcagaac aacatgtgcg cgcgcattcg gggcccgccag 1440
 cacctgtgc agctgaccgt gtggggcatc aagcagctgc aggcccgcat cctggccgtg 1500
 gagcgctacc tgaaggacca gcaatgtgcg ggcacatctggg gtcgcagccg caagctgatc 1560
 tgcaccacca cccgtgcctgtc gaacagcagc tggagcaaca agagcctgac cgagatctgg 1620
 gacaacatga cctggatggaa gtggggcgccg gagatcgccca actacacccg cctgatctac 1680
 aacctgtatcg agatcgccccca gaaccagcag gagaagaacg agcaggagct gctggagctg 1740
 gacaagtggg ccagcctgtg gaactgggtt gacatccca actggctgtg gtacatccgc 1800
 atcttcatca tgatcggtggg cggcctgtatc ggccctgcgca tgcgttccgc cgtgtgagc 1860
 atcgtgaacc gctgtgcgc gggctacagc cccatcagcc tgcagacccg cctggccgc 1920
 cagcgcggcc cccgaccgccc cgaggccatc gaggaggagg gcccggagccg cgaccgcgac 1980
 cgcagaacc gctgtggcga cggcctgtg ggccctgtatc gggacacct ggcgcggctg 2040
 tgccttca gctaccaccc cctgcgcgac ctgcgtgtatc tgcgtggccg catgtggag 2100
 ctgctggcc gccgcggctg ggaggccctg aagtactggt ggaacactgt gcaactgtgg 2160
 agccaggagc tgaagagcag cccgtgagc ctgttcaacg ccacccgcat cccgtggcc 2220
 gagggcaccg accgcattatcg ctagatgtgc cagcgcattt tccgcgcgt gatccacatc 2280
 ccccccgcga tccggccaggcc cctggagccg gccctgtgt aagatatacg atcccttaga 2340
 gaattccgc ccccccccccc ccccccctt cccctcccccc cccctaacgt tactggccga 2400
 agccgcttgg aataaggccg gtgtgcgttt gtcttatatgt tattttccac catattggcc 2460
 tcttttggca atgtggggc cccggaaacctt ggccctgtct tcttgcgag catttcttagg 2520
 ggtctttccc ctctcgccaa aggaatgtcaaa ggtctgttga atgtcgtaa ggaagcagtt 2580
 cctcttggaaat ctcttgcag acaaacaacg tctgtgcga ccctttgcag gcagcggAAC 2640
 ccccccacccgc ggcacagggtg cctctgcggc caaaagccac gtgtataaga tacacctgca 2700
 aaggcggcac aaccccaagtgc ccacgttgcg agtggatag ttgtggaaag agtcaaatgg 2760
 ctctcttcaaa gctgtattcaaa caaggggtgc aagatgtccc agaaggatacc ccattgtatg 2820
 ggatctgtatc tggggccctcg gtgcacatgc ttatcatgt tttatgtcgag gttaaaaaaaa 2880
 cgtcttagcc ccccgaaacca cggggacgtg gtttccctt gaaaaacacg ataataccat 2940
 gggcgccccgc gccagcgtgc tgagcggccgg cgagctggac aagtgggaga agatccgcct 3000
 ggcgcggccgc ggcagaagaaga agtacaagct gaagcacatc gtgtggccca gccgcgagct 3060
 ggagcgcttc gccgtgaacc cccgcctgtt ggagaccagc gagggtcgcc gccagatccct 3120
 gggccagctg cagcccaaggcc tgcagaccgg cagcgaggag ctgcgcagcc tgcgttccac 3180
 cgtggccacc ctgtactgcg tgcaccaggcc catcgacgtc aaggacacca aggaggccct 3240
 ggagaagatc gaggaggaggc agaacaatgc caagaagaag gcccagccgg cccgcggccgc 3300
 cggccggcacc ggcacacgc gccagggtg ccagaactac cccatcgatc agaacactgca 3360
 gggccagatg gtgcaccagg ccatcaggccc cccgaccctt aacgcctggg tgaaggtgg 3420
 ggaggagaag gctttcagcc cccgagggtat cccatgttc agcgcctgtg ggcaggccgc 3480
 ccccccccaag gacctgaaca cgtatgttgc caccgtggc ggcacccagg cccgcattgca 3540
 gatgtgttgcg gagaccatca acgaggaggc cccgcgttgc gaccgcgtgc accccctgtca 3600
 cggccggccccc atcgccccccg gccagatgcg cgagccccgc ggcagcgcaca tcggccggcac 3660

caccagcacc	ctgcaggagc	agatcggtcg	gatgaccaac	aacccccc	tccccgtggg	3720
cgagatctac	aagcggtgga	tcatcctggg	cctgaacaag	atcggtcgga	tgtacagccc	3780
caccagcatc	ctggacatcc	gccaggccc	caaggagccc	ttcccgact	acgtggaccg	3840
cttctacaag	accctgcg	ctgagcaggc	cagccaggac	gtgaagaact	ggatgaccga	3900
gaccctgctg	gtgcagaacg	ccaacccga	ctgcaagacc	atcctgaagg	ctctcgcccc	3960
cgcgccacc	ctggaggaga	tgatgaccgc	ctgcccaggc	gtggggcgcc	ccggccacaa	4020
ggcccgctg	ctggccgagg	cgatgagcca	ggtgacgaaac	ccggcgacca	tcatgatgca	4080
gcgcggcaac	ttccgcAAC	agcggaaagac	cgtcaagtgc	ttaaactgcg	gcaaggaggg	4140
ccacaccgccc	aggaactgcc	gcgcccccc	caagaaggcc	tgttggcgct	gcggcccgaa	4200
gggccaccag	atgaaggact	gcaccgagcg	ccaggccaaac	ttcttgggca	agatctggcc	4260
cagctacaag	ggccggcccc	gcaacttct	gcagagccgc	cccgagccca	ccgccccccc	4320
cgaggagagc	ttccgcTTcg	gcgaggagaa	gaccacccccc	agccagaagc	aggagcccat	4380
cgacaaggag	ctgtacccccc	tgaccagct	gcgcagcctg	ttcggcaacg	accccagcag	4440
ccagtaagaa	ttcagactcg	aqcaagtcta	qa			4472

<210> 76

<211> 4608

<212> DNA

<212> DNA

<220>

<223> Description of Artificial Sequence:

bp160_modSE162_dLV2_gar_modSE2

<400> 76

gaatccggca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtctcg tttcgccccag cgccgtggag aagctgtggg tgaccgtgtta ctacggcgtg 120
cccgtgtgga aggaggccac caccacccctg ttctggtgttca ggcacgccaa ggcctacgac 180
accgagggtgc aacaactgtg ggccacccac gcctgtgtgc ccaccgacc caaccccccag 240
gagatcgtdc tggagaacgt gaccgagaac ttcaacatgt ggaaagaacaa catggtgtag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
accccccgtg gcgtgaccct gcactgcacc aacctgaaga acgcccacca cccaagagc 420
agcaactgtga aggagatggg ccggccggag atcaagaact gcagcttcaa ggtggggcgcc 480
ggcaagctga tcaactgcaa caccagctg atcaccctagg cctgccccaa ggtgagcttc 540
gagcccatcc ccatccacta ctgcggccccc gccggcttcg ccatactgaa gtgcacacgac 600
aagaagttca acggcagcgg cccctgcacc aacgtgagca ccgtgcagtg cacccacggc 660
atccgccccg tggtagacac ccagctgtg ctgaacggca gcctggccga ggaggggctg 720
gtgatccgca gcgagaactt caccgacaac gccaagacca tcafcgtgca gctgaaggag 780
agcgtggaga tcaactgcac ccggcccaac aacaacaccc gcaagagcat caccatcgcc 840
cccgcccgcc cttctacgc caccggcgcac atcatcgccg acatccgcac ggccactgac 900
aacatcagcg gcgagaagtg gaacaacaccc ctgaagcaga tcgtgacccaa gtcgcaggcc 960
cagttcgca acaagaccat cgtgttcaag cagagcagcg gcggcgacc cggatctgt 1020
atgcacagct tcaactgcgg cgccgagttc ttctactgca acagcaccca gctgttcaac 1080
agcacctgtt acaacaccat ccggcccaac aacaccaacg gcaccatcac cctgcccgtc 1140
cgcatcaagc agatcatcaa ccgtgtggcag gaggtggcga aggccatgta cggccccccc 1200
atccgccccg agatccgctg cagcagcaac atcaccggcc tgctgctgac cggcgaaggc 1260
ggcaaggaga tcagcaacac caccgagatc ttccggccccc gcggcgccga catgcgcgac 1320
aactggcgca gcgagctgtt caagtacaag gtggtagaaa tcgagccctt gggcgtggcc 1380
cccaccaagg ccaagcggccg cgtgggtgcag cgcgagaagc gcgcgtgac cctggggcc 1440
atgttccctgg gcttccctggg cggccggccgc agcaccatgg gcgcggcccg cctgaccctg 1500
accgtgcagg cccggccagct gctgagcggc atcgtgcagc agcagaacaa cctgctgcgc 1560
gccccatcgagg cccagcagca cctgtgtgcag ctgaccgtgtt ggggcatcaa gcagctgcag 1620
gccccgggtgc tggccgtggg ggcgttacccgt aaggaccagc agctgctggg catctgggc 1680
tgtcagcggca agctgatctg caccacccggc gtgcctggaa acgcctgtgtc gagcaacaag 1740
agcctggacc agatctggaa caacatgacc tggatggagt gggagcgcga gatcgacaac 1800
tacaccaacc tgatctacac cctgtatcgag gagagccaga accagcagga gaagaacgag 1860
caggagctgc tggagctggg caagtggggc agcctgtggaa actgggttcga catcagcaag 1920
tggctgtgtt acatcaagat cttcatcatg atcgtggccgg gcctgggtggg cctgcgcattc 1980
gtgttccaccc tgctgagcat cgtgaaccgc gtgcggccagg gtcacagccc cctqaqcttc 2040

cagacccgct tccccggccc ccgcggccccc gaccgcggcc agggcatcga ggaggagggc 2100
 ggcgagcgcc accgcgaccg cagcagccccc ctgggtcacg gcctgctggc cctgatctgg 2160
 gacgacctgc gcagcctgtg cctgttcage taccaccggc tgcgccacct gatccctgatc 2220
 gccggccgca tcgtggagct gctggccgc cgccggctggg aggccctgaa gtactggggc 2280
 aacctgtgc agtactggat ccaggagctg aagaacagcg ccgtgagccgt gttcgacgcc 2340
 atcgccatcg cctgtggccga gggcaccgc acgcacatcg aggtggccca ggcgcattggc 2400
 cgccgccttcc tgcacatccc cggccgcate cgccagggtc tgcgcgcgc cctgtgtaa 2460
 ctcgagcaag tctagagaat tccggccccc ccccccccccc ccctctccct ccccccccccc 2520
 taacgttaact ggcgaagcc gcttggataa aggcgggtgt gctttgtct atatgttatt 2580
 ttccaccata ttgcgtctt ttggcaatgt gagggccccc aaacctggcc ctgtcttctt 2640
 gacgacatt cctagggttc ttccctctt cgcggaaaggaa atgcaagggtc tggtgaatgt 2700
 cgtgaaggaa gcagttcctc tggaaagttc ttgaagacaa acaacgtctg taggcaccc 2760
 ttgcaggcag cggaaacccccc cacctggcga caggtgcetc tgccggccaa agccacgtgt 2820
 ataagataca cctgcggaaagg cggcacaacc ccagtgcac gttgtgagtt ggatagtgt 2880
 ggaaagagtc aaatggctct cctcaagcgat attcaacaag gggctgaagg atgcccagaa 2940
 ggtaccccat tgtatggat ctgatctggg gcctcggtgc acatgttta catgtgttta 3000
 gtcgaggtta aaaaaacgtc taggcccccc gaaccacggg gacgtgggtt tccttgaaa 3060
 aacacgataa taccatgggc gcccggccca gctgtgtgag cggccggcag ctggacaagt 3120
 gggagaagat cgcgcgtgc cccggccggca agaagaagta caagctgaag cacatcggt 3180
 gggccagccg cgagctggag cggttcggcc tgaacccccc cctgtgtgag accagcgagg 3240
 gctgcggcca gatcctgggc cagctgcage ccagcctgca gaccggcagc gaggagctgc 3300
 gcagcctgtca caacaccgtg gcccacccgt actgcgtgca ccagcgcac gacgtcaagg 3360
 acaccaagga gggccctggag aagatcgagg aggagcagaa caagtccaa aagaaggccc 3420
 agcaggccgc cgccgcgcgc ggcaccggca acagcagcca ggtgagccag aactacccca 3480
 tcgtgcagaa cctgcaggcgc cagatgtgc accaggccat cagccccccg accctgaacg 3540
 cctgggtgaa ggtgggtggag gagaaggcct tcagccccca ggtgatcccc atgttcagcg 3600
 ccctgagcga gggccgcacc ccccagggacc tgaacacgat gttgaacacc gttggccggcc 3660
 accagggccgc catgcagatg ctgaaggaga ccatcaacgaa ggaggccgc gatggggacc 3720
 gctgtgcaccc cgtgcacgc gggcccatcg ccccccggcca gatgcgcgag ccccgccgca 3780
 gcgacatcgc cggcaccacc accgcacccgtc aggagcagat cggctggatg accaacaacc 3840
 ccccccattccc cgtggggcag atctacaagc ggtggatcat cctgggcctg aacaagatcg 3900
 tgcggatgtc cagccccacc agcatctgg acatccggca gggcccccaag gagcccttcc 3960
 gcgactacgt ggaccgcctt tacaagaccc tgcgcgtca gcaggccagc caggacgtga 4020
 agaactggat gaccgagacc ctgctggtgc agaacgccaa ccccgactgc aagaccatcc 4080
 tgaagctct cggcccccgcg gcccacccgtt aggagatgt gaccgcctgc cagggcgtgg 4140
 gcgcccccgg ccacaaggcc cgcgtgttcc cggaggcgtt gagccaggtg acgaacccgg 4200
 cgaccatcat gatgcagcgc ggcaacttcc gcaaccagcg gaagacgctc aagtgttca 4260
 actgcggcaa ggagggccac accgcacccggactgc ccccccgcgaa aagggtgtct 4320
 ggcgcgtgcgg cccgcggaggc caccagatga aggactgcac cgagcgcac gccaacttcc 4380
 tggggcaagat ctggcccgac tacaaggccc gccccggcaa cttccctgcag agcccccgg 4440
 agcccaaccgc ccccccccgag gagagcttcc gcttcggcga ggagaagacc accccccagcc 4500
 agaaggcagga gcccacatcgac aaggagctgt accccctgac cagcctgcgc agcctgttgc 4560
 gcaacgaccc cagcagccag taagaattca gactcgagca agtctaga 4608

<210> 77

<211> 1680

<212> DNA

<213> Human immunodeficiency virus

<400> 77

cccat tagtc ctattgaaac tgttaccagta aaattaaagc caggaatgga tggccaaaa 60
 gttaaagcaat ggcattgtac agaagaaaaaa ataaaagcat tagtagagat atgtacagaa 120
 atggaaaagg aaggggaaaaat ttcaaaaaattt gggcctgaaa atccatacaa tactccagta 180
 tttgtataaa aaaaaaaaaa cagtactaaa tggagaaaaac tagtagattt cagagaactt 240
 aataaaaaggaa ctcaagactt ctgggaagtt cagtttagaa taccacaccc cgccagggtt 300
 aaaaaagaaaa aatcagtaac agtattggat gtgggtgtatc catacttttc agtcccctt 360
 gataaaagact ttagaaagta tactgcattt accataccca gtataaaca tgagacacca 420
 gggatttagat atcagtacaa tgtgtgcac cagggatgga aaggatcacc agcaatattc 480
 caaagttagca tgacaaaaat ctttagagcc ttttagaaaaac agaatccaga catagtttac 540

tatcaatacata tggatgattt gtatgttagga tctgacttag aaataggcca gcatagaaca 600
 aaaatagagg aactgagaca gcatctgtt aggtggggat ttaccacacc agacaaaaaa 660
 catcagaaaag aacctccatt cctttggatg ggttatgaac tccatcccta taaatggaca 720
 gtacagccta taatgctgcc agaaaaagac agtggactg tcaatgacat acagaagtta 780
 gtggggaaaat tgaattgggc aagtcaattat tatgcagggg ttaaagtaaa gcagttatgt 840
 aaactcccta gaggaaccaa agcactaaca gaagtaatac cactaacaga agaagcagag 900
 ctagaactgg cagaaaacag ggagattcta aaagaaccag tacatgaagt atattatgac 960
 ccatcaaaaag acttagtagc agaaaatacag aagcaggccc aaggccaatg gacatatcaa 1020
 atttatcaag agccatttaa aaatctgaaa acaggaaagt atgcaaggat gagggtgcc 1080
 cacactaatg atgtaaaaca gttAACAGAG gcagtgcataa aagtatccac agaaagcata 1140
 gtaatatggg gaaagattcc taaattttaa ctacccatac aaaaggaaac atgggaagca 1200
 tgggtggatgg agtattggca agtacactgg attcctgagt gggagttgt caataccct 1260
 cccttagtga aattatggta ccagtttagag aaagaacccc tagtaggac agaaacttgc 1320
 tatgttagatg gggcagctaa tagggagact aaatttagaa aagcaggata tgtaactgac 1380
 agaggaagac aaaaagttgt ctccatagct gacacaacaa atcagaagac tgaattacaa 1440
 gcaattcatac tagctttca ggattccggta tttagaagttt acatgttaac agactcacaa 1500
 tatgcatttag gaatcattca agcacaacca gataagagt aatcagatg agtcagtcaa 1560
 ataatacagac agttaataaa aaaggaaaag gtctacctgg catgggtacc agcacacaaa 1620
 ggaattggag gaaatgaaca agtagataaa ttagtcagtg ctggaaatcg gaaagtacta 1680

<210> 78

<211> 1865

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: GP1

<400> 78

gtcgacgcca ccatgggcgc cggcgccagc gtgctgagcg gggcgagct ggacaagtgg 60
 gagaagatcc gcctgcgccc cggcgccaaag aagaagtaca agctgaagca catctgtgtgg 120
 gccagcccg agctggagcg cttegcgtg aacccccggcc tgctggagac cagcgaggcc 180
 tgccgcccaga tcctgggcca gctgcagcccc agcctgcaga cccgcagcga gggactgcgc 240
 agcctgtaca acaccgtggc caccctgtac tgcgtgcacc agcgcatacg cgtcaaggac 300
 accaaggagg ccctggagaa gatcgaggag gaggcagaaca agtccaagaa gaaggcccag 360
 caggccgccc cccgcggccgg caccggcaac agcagccagg tgagccagaa ctacccatc 420
 gtgcagaacc tgcagggcca gatgggtcac caggccatca gcccccgac cctgaacgccc 480
 tgggtgaagg tgggtggagga gaaggccttc agccccgagg tgatccccat gttcagcgcc 540
 ctgagcgagg gcccaccccc ccaggacactg aacacgtgt tgaacacccgt gggcgccac 600
 caggccgcca tgcagatgct gaaggagacc atcaacgagg aggccgcccgt gttggaccgc 660
 gtgcaccccg tgcacgcggg ccccatcgcc cccggccaga tgccgcagcc cccggccagc 720
 gacatcgccg gcaaccaccag caccctgcag gagcagatcg gctggatgac caacaacccc 780
 cccatccccg tggcgagat ctacaagcgg tggatcatcc tgggctgaa caagatcg 840
 cggatgtaca gccccaccag catctggac atccgcagg gcccccaagga gcccctccgc 900
 gactacgtgg accgcttcta caagaccctg cgcgcgtgac agggcagccca ggacgtgaag 960
 aactggatga ccgagaccct gctgggtcac aacgcacaacc cgcactgcac gaccatcctg 1020
 aaggctctcg gcccccgccg caccctggag gagatgtga cgcctgcac gggctgtggc 1080
 gggcccgcc acaaggccccg cgtgtggcc gaggcgtga gcccagggtac gaaccggcg 1140
 accatcatga tgcagcgccg caacttccgc aaccagcggc agaccgtcaa gtgcttcaac 1200
 tgccgcagg agggccacac cgccaggaaac tgccgcgc cccgcaga gggctgtggc 1260
 cgctgcggcc gccaaggaca ccaaattgaaa gattgcactg agagacagc taattttta 1320
 gggaaatctt ggccttcata caagggaaagg ccaggaaatt ttcttcagag cagaccagag 1380
 ccaacagccc caccagaaga gagcttcagg tttggggagg agaaaacaac tcccttcag 1440
 aagcaggagc cgatagacaa ggaactgtat ccttaactt ccctcagatc actcttggc 1500
 aacgaccctt cgtcacagta aggtcgccg gccagctcaa ggaggcgcgt ctcgcacaccg 1560
 gccgcgcacga caccgtgtc gaggagatga acctgcccgg caagtggaaag cccaaagatga 1620
 tcggcgggat cgggggcttc atcaaggatgc ggcagtacga ccagatcccc gtggagatct 1680
 gccgcacaa gcccacatcgcc accgtgtgg tggccccac ccccggtgaac atcatcgcc 1740
 gcaacctgct gacccagatc ggctgcaccc tgaacttccc catcagcccc atcgagacgg 1800

tgcccggtgaa gctgaagccg gggatggacg gccccaagggt caagcagtgg cccctgttaag 1860
aattc 1865

<210> 79
<211> 1865
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: GP2

<400> 79
gtcgacgcca ccatgggcgc cgcgcgcgcgtgctgagcg gggcgagct ggacaagtgg 60
gagaagatcc gcctgcgcgc cggcgcaag aagaagtaca agctgaagca catcggtgtgg 120
gccagccgcg agctggagcg ctgcgcgtg aaccccgccc tgctggagac cagcgaggc 180
tgccggccaga tcctgggcca gctgcagccc agcgtcgaga cccgcagcga ggagctgcgc 240
agcctgtaca acaccgtggc caccctgtac tgctgcacc agcgcatacg cgtcaaggac 300
accaaggagg ccctggagaa gatcgaggag gacgagaaca agtccaagaa gaaggcccag 360
caggccgcgc cccgcgcgcg caccggcaac agcgcagg tgagccagaa ctacccatc 420
gtgcagaacc tgcaggggcca gatggtcac caggccatca gccccgcac cctgaacgccc 480
tgggtgaagg tgggtggagga gaaggccttc agcccccggg tgatccccat gttagcgc 540
ctgagcgagg ggcgcacccc ccaggacactg aacacgtgt tgaacacccgt gggcgccac 600
caggccgcgc tgcagatgtc gaaggagacc atcaacgagg agggccgcga gtgggaccgc 660
gtgcaccccg tgcacgcgcgc ccccatcgcc cccggccaga tgccgcagcc cccgcgcgc 720
gacatcgccg gcaccaccag caccctgcag gagcagatcg gctggatgac caacaacccc 780
cccatccccg tgggcgagat ctacaacggg tggatcatcc tgggcctgaa caagatcg 840
cgatgtaca gcccccacca catcctggac atccgcagg gcccccaagga gcccctccgc 900
gactacgtgg accgcttcta caagaccctg cgcgtgagc aggccagcca ggacgtgaag 960
aactggatga cgcgagaccct gctggtcag aacgcaccc cccactgcaaa gaccatcctg 1020
aaggctctcg gcccccgccgc caccctggag gagatgtga cgcctgcca gggcggtggc 1080
ggccccggcc acaaggcccg cgtgctggcc gaggcgtga gccaggtgac gaacccggcg 1140
accatcatga tgcagcgccg caacttccgc aaccagcgga agaccgtcaa gtgttcaac 1200
tgcggcaagg agggccacac cgccaggAACAC tgccgcgcgc cccgcaagaa gggctgctgg 1260
cgctcgccgc gccaaggaca ccaaattaaaa gattgcactg agagacaggc taattttta 1320
gggaagatct ggccttccta caagggaaagg ccagggaaatt ttcttcagag cagaccagag 1380
ccaaacagccc caccagaaga gagcttcagg tttggggagg agaaaacaac tccctctcag 1440
aagcaggagc cgatagacaa ggaactgttat ctttaactt ccctcagatc actctttggc 1500
aacgaccct cgtcacagta aggatcgggg ggcaactcaa ggaagcgctg ctcgatacag 1560
gagcagatga tacagtatta gaagaaatga atttgcagg aaaatggaaa ccaaaaatga 1620
tagggggat cgggggcttc atcaaggtga ggcgtacga ccagataacct gtggaaatct 1680
gtggacataa agctataggt acagtattag taggacctac acctgtcaac ataattggaa 1740
gaaatctgtt gacccagatc ggctgcaccc tgaacttccc catcagccct attgagacgg 1800
tgcccggtgaa gttgaagccg gggatggacg gcccccaagggt caagcaatgg ccattgttaag 1860
aattc 1865

<210> 80
<211> 2305
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
FS(+) .proinact.RTopt.YM

<400> 80
gcggccgcga aggacaccaa atgaaagatt gcactgagag acaggctaatttttaggga 60
agatctggcc ttccataag ggaaggccag ggaattttct tcagagcaga ccagagccaa 120
cagccccacc agaagagagc ttcaaggtttgggaggagaa aacaactccc tctcagaagc 180
aggagccgat agacaaaggaa ctgtatcctt taacttccct cagatcactc tttggcaacg 240

acccctcgtaacaataagga tcggggggca actcaaggaa gcgctgctcg atacaggagc 300
 agatgataca gtattagaag aaatgaattt gccaggaaaa tggaaaccaa aaatgatagg 360
 ggggatcggg gccttcatca aggtgaggca gtacgaccag atacctgttag aaatctgtgg 420
 acataaaagct ataggtacag tattatggg acctacacct gtcaacataa ttgaaagaaa 480
 tctgttgacc cagatcggct gcacctgaa ctccccatc agccctattt agacggtgcc 540
 cgtgaagttt aagccgggaa tggacggccc caaggtcaag caatggccat tgaccgagga 600
 gaagatcaag gccttgggg agatctgcac cgagatgggg aaggaggggca agatcagcaa 660
 gatcggcccc gagaacccct acaacacccc cgtgttcggc atcaagaaga aggacagcac 720
 caagtggcgc aagctgggg acttccggca gctgaacaag cgacccagg acttctggg 780
 ggtgeagctg ggcatcccc accccggccg cctgaagaag aagaagagcg tgaccgtgct 840
 ggacgtgggc gacgcctact tcagegtgcc cctggacaag gacttccgca agtacaccgc 900
 cttcaccatc cccagcatca acaacgagac cccggcatc cgctaccagt acaacgtgct 960
 gccccaggggc tggaaaggca gccccggcat ctccagagg agcatgacca agatccctgga 1020
 gcccctccgc aagcagaacc ccgcacatcg gatctaccag gccccctgt acgtggggcag 1080
 cgacotggag atcggccagc accgcaccaa gatcgaggag ctgegcacgc acctgctgct 1140
 ctggggcttc accacccccc acaagaagca ccagaaggag ccccccttc tgtggatggg 1200
 ctacgagctg caccggaca agtggaccgt gcagcccatc atgtgcggc agaaggacag 1260
 ctggaccgtg aacgacatcc agaagctggt gggcaagctg aactggggca gccagatcta 1320
 cgccggcatc aagggtgaagc agctgtcaaa gctgtgcgc ggcaccaagg ccctgaccga 1380
 ggtgatcccc ctgaccgagg aggccgagct ggagctggcc gagaaccggc agatccctgaa 1440
 ggagcccggtg caccgggtt actacgaccc cagaaggac ctggggccg agatcccgaa 1500
 gcagggccag ggccagtggc cctaccagat ctaccaggag cccttcaaga acctgaagac 1560
 cggcaagtac gcggcgatgc gcggcgcccc caccacgc gatcgaggag tgaccgagc 1620
 cgtcagaag gtgagcaccc agagcatcgat gatctggggc aagatcccca agtcaagct 1680
 gccccatccag aaggagaccc gggaggctg gtggatggg tactggcagg ccacctggat 1740
 ccccgagtgg gagttcgtga acacccccc cctggtaag ctgtggtacc agctggagaa 1800
 ggagcccatc gtggggcgcc agaccttcta cttggacggc gccgccaacc gcgagaccaa 1860
 gctgggcaag gcggctacg tgaccgaccg gggccggcag aaggtggtga gcatcgccga 1920
 caccaccaac cagaagaccc agctgcaggc catccacctg gccctgcagg acagccgcct 1980
 ggaggtgaac atcgtgaccg acagccagta cgcctggc atcatccagg cccagcccg 2040
 caagagcgag agcgagctgg tgagccagat catcgaggc ctgatcaaga aggagaaggt 2100
 gtacctggcc tgggtggccccc cccacaaggg catcgccggc aacgagcagg tggacaagct 2160
 ggtgagcgcc ggcaccccgca aggtgtgtt cctgaacggc atcgatggcg gcatcgatgt 2220
 ctaccagtac atggacgacc tgtacgtggg cagcggcgcc cctaggatcg attaaaagct 2280
 tcccgccgtc agcaccgggtg aattc 2305

<210> 81

<211> 2299

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
FS(+) .proinact.RTopt.YMWM

<400> 81

gcggccgcga aggacaccaa atgaaagatt gcactgagag acaggctaatttttaggga 60
 agatctggcc ttccatcaag ggaaggccag ggaattttct tcagagcaga ccagagccaa 120
 cagccccacc agaagagagc ttccagggtt gggaggagaa aacaactccc tctcagaagc 180
 aggagccgat agacaaggaa ctgttatccct taacttccct cagatcactc tttggcaacg 240
 acccctcgtaacaataagga tcggggggca actcaaggaa ggcgtgctcg atacaggagc 300
 agatgataca gtattagaag aaatgaattt gccaggaaaa tggaaaccaa aaatgatagg 360
 ggggatcggg gccttcatca aggtgaggca gtacgaccag atacctgttag aaatctgtgg 420
 acataaaagct ataggtacag tattatggg acctacacct gtcaacataa ttgaaagaaa 480
 tctgttgacc cagatcggct gcacctgaa ctccccatc agccctattt agacggtgcc 540
 cgtgaagttt aagccggggaa tggacggccc caaggtcaag caatggccat tgaccgagga 600
 gaagatcaag gccttgggg agatctgcac cgagatgggg aaggaggggca agatcagcaa 660
 gatcggcccc gagaacccct acaacacccc cgtgttcggc atcaagaaga aggacagcac 720
 caagtggcgc aagctgggg acttccggca gctgaacaag cgacccagg acttctggg 780

ggtgcagctg ggcatcccc accccggcgg cctgaagaag aagaagagcg tgaccgtgct 840
 ggacgtggc gacgcctact tcagcgtgcc cctggacaag gacttccgca agtacaccgc 900
 cttcaccatc cccagcatca acaaaggagac ccccgccatc cgctaccagt acaacgtgct 960
 gccccagggc tggaaaggca gccccccat cttccagagc agcatgacca agatccctgga 1020
 gcccctccgc aacgagaacc cccgacatcgat gatctaccag gccccccctgt acgtggcag 1080
 cgacctggag atcggccagc accgcaccaa gatcgaggag ctgcgcacgc acctgctgct 1140
 ctggggcttc accacccccc acaagaagca ccagaaggag ccccccccttc tgcccatcga 1200
 gtcgcacccc gacaagggtga cccgtgcaccc catcatgtcg cccgagaagg acagctggac 1260
 cgtgaacgac atccagaagg tggtggcaa gtcgaactgg gccagccaga tctacgcccgg 1320
 catcaaggtg aacgagctgt gcaagctgct gcgcggcacc aaggccctga ccgaggttat 1380
 cccccctgacc gaggaggccg agctggagct ggccgagaac cgcgagatcc tgaaggagcc 1440
 cgtgcacgac gtgtactacg accccagcaa ggacctggtg gccgagatcc agaagcaggg 1500
 ccaggcccg tgacacctacc agatctacca ggagcccttc aagaacctga agaccggcaa 1560
 gtacgcccgc atgcgcggcg cccacaccaa cgacgtgaag cagctgaccg aggccgtgca 1620
 gaaggtgagc accgagagca tcgtgatctg gggcaagatc cccaagttca agctgcccatt 1680
 ccagaaggag acctgggagg cctgggttat ggagttactgg cagggccactt ggatccccga 1740
 gtgggagttc gtaaacaccc ccccccttgtt aaagctgtgg taccagctgg agaaggagcc 1800
 catcggtggc gccgagaccc tctacgtgga cggcggcc aaccgcgaga ccaagctggg 1860
 caaggccgc tacgtgaccg accggggccg gcagaagggt gtgagcatcg cgcacaccac 1920
 caaccagaag accgagctgc aggccatcca cctggccctg caggacagcg gcctggaggt 1980
 gaacatcgtg accgacagcc agtacgccc gggcatcata caggcccgcc cgcacaagag 2040
 cgagagcgag ctggtgagcc agatcatcga gcagctgatc aagaaggaga aggtgtaccc 2100
 ggcctgggtg cccgcccaca agggcatcgg cggcaacgag caggtggaca agctggtag 2160
 cggccgcate cgcaagggtgc tggctctgaa cggcatcgat ggcggcatcg tgatctacca 2220
 gtacatggac gacctgtacg tggcagcgg cggccctagg atcgattaaa agcttccgg 2280
 ggcttagcacc ggtgaattc 2299

<210> 82

<211> 2306

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

FS(-).protmod.RTopt.YM

<400> 82

gcgccgcga aggacaccaa atgaaagatt gcactgagag acaggcta at tctccgcg 60
 aggacctggc cttcctgcag ggcaaggccc gcgagttcg cagcgagcag acccgccca 120
 acagccccac cccgcgcgag ctgcagggtt gggcgccga gaacaacagc ctgagcgagg 180
 cccgcgcga cccgcaggcc accgtgagct tcaacttccc ccagatcacc ctgtggcagc 240
 gcccccttgtt gaccatcagg atcggcggcc agtcaagga ggcgctgctc gacaccggcg 300
 cccgacgacac ctgtgctggag gagatgaacc tggccggcaat gtgaaagccc aagatgtatcg 360
 gcgggatcgg gggcttcatc aaggtgcggc agtacgacca gatccccctg gagatctgcg 420
 gccacaaggc catccgcacc gtgctggtg gccccccccc cgtgaacatc atcgccgc 480
 acctgctgac ccagatcgcc tgcaccctga acttccccc cagccccatc gagacgggtgc 540
 ccgtgaagct gaagccgggg atggacgcgc ccaaggtcaa gcagtggccc ctgaccgagg 600
 agaagatcaa gggcttggtg gagatctgc acgagatggc gaaggaggcc aagatcagca 660
 agatcgccccc cgagaaccccc tacaacaccc ccgtgttcgc catcaagaag aaggacagca 720
 ccaagtggcg caagctggtg gacttccgcg agtcaacaa gcgcaccacag gacttctggg 780
 aggtgcagct gggcatcccc caccggccgc gcctgaaagaa gaagaagagc gtgaccgtgc 840
 tggacgtggc cgacgcctac ttctgcgtgc ccctggacaa ggacttccgc aagtacaccg 900
 cttcaccat ccccgccatc aacaacgaga ccccgccat cccgtaccacg tacaacgtgc 960
 tggcccccggg ctggaaaggcc agccccccca tctttccagag cagcatgacc aagatccctgg 1020
 agccctccg caagcagaac cccgacatcg tgatctacca gggcccccctg tacgtggc 1080
 gcgacctggc gatcgcccg caccgcacca agatcgagga gtcgcgcacg cacctgctgc 1140
 gctggggctt caccacccccc gacaagaagc accagaagga gcccccccttc ctgtggatgg 1200
 gctacgagct gcaccccgac aagtggaccc tgcaagccat catgctgccc gagaaggaca 1260
 gctggaccgt gaacgacatc cagaagctgg tggcagccatc catgctgccc gagaaggaca 1320

acgcggccat caaggtgaag cagctgtca agctgctgcg cggcacaag gccctgaccg 1380
 aggtgatccc cctgaccgag gaggccgagc tggagctggc cgagaaccgc gagatcctga 1440
 aggagcccggt gcacgaggt tactacgacc ccagcaagga cctggtgcc gagatccaga 1500
 agcagggcca gggccagtgg acctaccaga tctaccaggaa gcccctaag aacctgaaga 1560
 cccgcaagta cgccccatcg cgccgcggcc acaccaacga cgtgaagcag ctgaccgagg 1620
 ccgtcagaa ggtgagcacc gagagcatcg tgatctgggg caagatcccc aagtctcaagc 1680
 tgcccattcca gaaggagacc tgggagggct ggtggatggaa gtactggcag gccacctgga 1740
 tccccagtg ggagttcgta aacacccccc ccctggtcaa gctgtggtac cagctggaga 1800
 aggagcccat cgtgggcgcg gagacccctt acgtggacgg cggcccaac cgcgagacca 1860
 agctggcaaa ggccggctac gtgaccgacc gggccggca gaaggtggtg agcacatcgccg 1920
 acaccaccaa ccagaagacc gagctgcagg ccatccaccc ggcctgcag gacagcggcc 1980
 tggaggtgaa catcgtgacc gacagccagt acgcctggg catcateccag gcccagcccg 2040
 acaagagcga gagcggagctg gtgagccaga tcatcgagca gctgatcaag aaggagaagg 2100
 tgtacctggc ctgggtgccc gcccacaagg gcacatcgccg caacgagcag gtggacaagc 2160
 tggtagcgc cggcatccgc aaggtgtgt tcctgaacagg catcgatggc ggcacatcgta 2220
 tctaccagta catggacgac ctgtacgtgg gcagcggccg ccctaggatc gattaaaagc 2280
 ttccccgggc tagcacccgt gaattc 2306

<210> 83

<211> 2300

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

FS (-).protmod.RTopt.YMWM

<400> 83

gcggccgcga aggacaccaa atgaaagatt gcactgagag acaggctaatttctccgcg 60
 aggacctggc ctcctgcag ggcaaggccc gcgagttcag cagcgagcag acccgccca 120
 acagccccac ccgccgcgag ctgcagggtt gggccggcga gaacaacage ctgagcggagg 180
 cccggccgaa ccgcaggcacc accgtgagct tcaacttccc ccagatcacc ctgtggcagc 240
 gcccctggt gaccatcagg atcggccggcc agctcaagga ggcgctgtc gacaccggcg 300
 ccgacgacac cgtgctggag gagatgaacc tgcccgcaaa gtggaagccc aagatgtcg 360
 gcgggatcgg gggcttcata aaggtgcggc agtacgacca gatccccgtg gagatctgcg 420
 gcccacaaggc catcggcacc gtgctggtgg gcccacccccc cgtgaacatc atcggccgca 480
 acctgtgtac ccagatcgcc tgcacccctga acttccccat cagccccatc gagacgggtc 540
 ccgtgaagct gaagccgggg atggacggcc ccaaggtcaa gcagtggccc ctgaccgagg 600
 agaagatcaa ggccttggt gагатctгca cсgагатggaa гаaggагggc aагаттагca 660
 agatcggccc cgagaaccccc tacaacacccc cctgttgcg catcaagaag aaggacagca 720
 ccaagtggcg caagctggt gacttccggc agtgaacaa ggcaccccaag gacttctggg 780
 aggtgcagct gggatcccc caccggccg gcctgaagaa gaagaagagc gtgaccgtgc 840
 tggacgtggg cgacgcctac ttcaegctgc cctggacaa ggacttccgc aagtacacccg 900
 ccttcaccat ccccagcat aacaacgaga ccccccggcat ccgttaccag tacaacgtgc 960
 tggccctggg ctggaaaggc agccccccca tctttccagag cagcatgacc aagatctgg 1020
 agcccttccg caagcagaac cccgacatcg tgatctacca ggccccctt tacgtggca 1080
 ggcacccggaa gatcgccag caccgcacca agatcgagga gctgcggccag caccctgtgc 1140
 gctggggctt caccacccccc gacaagaaggc accagaaggaa gcccccccttc ctgccccatcg 1200
 agctgcaccc cgacaagtgg accgtgcaggcc ccatcatgt gccccggaaag gacagctgg 1260
 ccgtgaacga catccagaag ctggtggcga agtgaactg ggcaccccg atctacgcgg 1320
 gcatcaaggt gaagcagctg tgcaagctgc tgccggccac caaggccccctt accgagggtga 1380
 tccccctgac cgaggaggcc gagctggagc tggccggagaa ccgcggatc ctgaaggagc 1440
 ccgtgcacga ggtgtactac gaccccgacca aggacctggt ggcacccggatc cagaaggcagg 1500
 gcccaggccca gtggacccatc cagatctacc agggccccctt caagaacctg aagacccggca 1560
 agtaccccg catgcgcggc gcccacacca acgacgtgaa gcagctgacc gaggccgtgc 1620
 agaagggtgag caccgagagc atcgtatctt gggggcaagat ccccaagttc aagctgcggca 1680
 tccagaagga gaccctgggag gcctggtggaa tggagttactg gcaggccacc tggatccccg 1740
 agtgggagtt cgtgaacacc ccccccctgg tgaagctgtg gtaccagctg gagaaggagc 1800
 ccatcgtggg cgccgagacc ttctacgtgg acggccggcc caaccgcgag accaagctgg 1860

gcaaggccgg ctacgtgacc gaccggggcc ggcagaagggt ggtgagcata gcccacacca 1920
 ccaaccagaa gaccgagctg caggccatcc acctggccct gcaggacagc ggcctggagg 1980
 tgaacatcggt gaccgacagc cagtacgccc tgggcatcat ccaggcccag cccgacaaga 2040
 gcgagagcga gctggtgagc cagatcatcg agcagctgat caagaaggag aaggtgtacc 2100
 tggcttgggt gccccccac aaggcatcg gcccacacga gcagggtggac aagctggta 2160
 gcccggcat ccccaagggtg ctgttctga accgcatacg tggccgatc gtatctacc 2220
 agtacatgga cgaccgtac gtggcagcg gccccttag gatcgattaa aagttcccg 2280
 gggctagcac cggtaattc 2300

<210> 84

<211> 2312

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 FS (-).protmod.RTopt(+)

<400> 84

gcggccgcga aggacaccaa atgaaagatt gcactgagag acaggctaatt ttcttcgcg 60
 aggacctggc ttctctgcag ggcaaggccc gcggatctcg cagcgagcag acccgcgcca 120
 acagccccac cccgcgcgag ctgcagggtgt ggggcggcga gaacaacagc ctgagcgagg 180
 cccgcgcgca cccgcagggc accgtgagct tcaacttccc ccagatcacc ctgtggcgc 240
 gcccctggt gaccatcagg atcggcggcc agctcaagga ggcgtgtc gacaccggcg 300
 ccgacgacac cgtgctggag gagatgaacc tgcccgccaa gtggaaaggcc aagatgatcg 360
 gcgggatcgg gggcttcattc aaggtgcggc agtacgacca gatccccgtg gagatctgcg 420
 gccacaaggc catcgccacc gtgctgggg gccccaccccc cgtgaacatc atcggccgca 480
 acctgtgac ccagatcgcc tgcaccctga acttccccat cagccccat gagacgggtgc 540
 ccgtgaagct gaagccgggg atggacggcc ccaaggtaa gcagtggccc ctgaccggagg 600
 agaagatcaa ggccctgggt gagatctgcg cccgagatggc gaaggaggcc aagatcagca 660
 agatcgcccc cgagaacccc tacaacaccc cctgtttcgc catcaagaag aaggacagca 720
 ccaagtggcg caagctgggt gacttccgcg agctgaacaa gcgcaccccg gacttctggg 780
 aggtgcagct gggcatcccc caccggccg gcctgaagaa gaagaagagc gtgaccgtgc 840
 tggacgtggg cgacgcctac ttctcggtgc ccctggacaa ggacttcgc aagtacaccg 900
 ctttaccat ccccgacatc aacaacgaga ccccccgcatt cctgaccatc tacaacgtgc 960
 tgccccaggg ctggaaaggcc agccccgcct tcttccagag cagcatgacc aagatcctgg 1020
 agcccttccg caagcagaac cccgacatcg tgcattacca gtacatgac gacctgtacg 1080
 tggcagcga cttggagatc ggccagcacc gcaccaagat cgaggagctg cgccagcacc 1140
 tgctgcgtg gggcttcacc acccccgaca agaagcaccat gaaggaggccc cccttcctgt 1200
 ggtatgggcta cgagctgcac cccgacaatgg gacccgtgc gcccatcatg ctgcccggaga 1260
 aggacagctg gaccgtgaac gacatccaga agctgggtgg caagctgaac tggccagcc 1320
 agatctacgc cggcatcaag gtgaagcgc tgcgtcaagct gctgcgcggc accaaggccc 1380
 tgaccggaggt gatccccctcg accggaggagg cccgagctggc gctggccggag aaccggcaga 1440
 tcctgaagga gccccgtgcac gaggtgtact acgaccccgat caaggacctg tggccggaga 1500
 tccagaagca gggccaggcc cagtgacccctt accagatcta ccaggagccc ttcaagaacc 1560
 tgaagacccgg caagtacgc cccatgcgcg gcggccacac caacgacgtg aagcagctga 1620
 cccgaggccgt gcagaagggtg agcaccggaga gcatcgat ctggggcaag atccccaaatg 1680
 tcaagctgcc catccagaag gagacctggg aggctgggt gatggagatc tggcaggcca 1740
 cctggatcccc ctagtggggat ttctgtacca cccccccctt ggtgaagctg tggtaccagc 1800
 tggagaaggg accccatcgat ggcggccaga cttctacgt ggacggccgc gccaaccgcg 1860
 agaccaagct gggcaaggcc ggctacgtg cccgaccgggg cccggcagaag gtgggtgagca 1920
 tcggccgacac caccaaccag aagaccggac tgcaggccat ccacctgcgc ctgcaggaca 1980
 gcccggctggaa ggtgaacatc gtgaccgaca gccagatcg cctggccatc atccaggccc 2040
 agccccacaa gagcggagac gagctggta gccagatcat cgaggacgtg atcaagaagg 2100
 agaagggtgt a cttggccctgg gtggccccc acaaggccat cggccggcaac gagcagggtgg 2160
 acaagctgtt gggccggcc atccgcaagg tgcgttccat gacccgtatc gatggccgca 2220
 tcgttatca ccagtacatcg gacgaccctgt acgtggccag cggccggccctt aggatcgatt 2280
 aaaagttcc cggggcttagc accggtaat tc 2312

<210> 85
<211> 306
<212> DNA
<213> Human immunodeficiency virus

<400> 85
atggagccag tagatcctag attagagccc tggaaagcatt caggaagtca gcctaagact 60
gcttgtacaa attgttattt taaaaagtgt tgctttcatt gccaagtttg tttcataaca 120
aaaggcttag gcatctccta tggcaggaag aagcgagac agcgacgaaag agtcctcca 180
gacagtgggg ttcatcaagt ttctctacca aagcaaccccg ctccccagcc ccaaggggac 240
ccgcacaggcc cgaaggaaatc gaagaagaag gtggagagag agacagagac agatccagtc 300
cattag 306

<210> 86
<211> 101
<212> PRT
<213> Human immunodeficiency virus

<400> 86
Met Glu Pro Val Asp Pro Arg Leu Glu Pro Trp Lys His Pro Gly Ser
1 5 10 15

Gln Pro Lys Thr Ala Cys Thr Asn Cys Tyr Cys Lys Lys Cys Cys Phe
20 25 30

His Cys Gln Val Cys Phe Ile Thr Lys Gly Leu Gly Ile Ser Tyr Gly
35 40 45

Arg Lys Lys Arg Arg Gln Arg Arg Arg Ala Pro Pro Asp Ser Glu Val
50 55 60

His Gln Val Ser Leu Pro Lys Gln Pro Ala Ser Gln Pro Gln Gly Asp
65 70 75 80

Pro Thr Gly Pro Lys Glu Ser Lys Lys Val Glu Arg Glu Thr Glu
85 90 95

Thr Asp Pro Val His
100

<210> 87
<211> 306
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: tat.SF162.opt

<400> 87
atggagcccg tggaccggcc cctggagccc tggaaagcacc ccggcagcca gccccagacc 60
gcctgcacca actgttactg caagaagtgc tgcttccact gccaggtgtg ctcatcacc 120
aaggccctgg gcatcagcta cggccgcaag aagcgccgac agcgccgccc cgcccccccc 180
gacagccgagg tgcaccagggt gagcctgccc aagcagcccg ccagccagcc ccagggcgac 240
cccaccggcc ccaaggagag caagaagaag gtggagcgac agaccgagac cgaccgggtg 300
cactag 306

<210> 88
<211> 306

<212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 tat.cys22.SF162.opt

<400> 88

```
atggagcccg tggacccccc cctggagccc tggaaaggcacc ccggcagcca gcccaagacc 60
gccggcacca actgctactg caagaagtgc tgcttccact gccaggtgtg cttcatcacc 120
aagggcctgg gcatcagcta cggccgcaag aagcggccgc agcgccggc cgcccccccc 180
gacagcgagg tgcaccaggat gagecctggcc aagcagcccg ccagccagcc ccagggcgac 240
cccacccggcc ccaaggagag caagaagaag gtggagcgcg agaccgagac cgacccctgtg 300
cacttag 306
```

<210> 89

<211> 168

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 tatamino.SF162.opt

<400> 89

```
atggagcccg tggacccccc cctggagccc tggaaaggcacc ccggcagcca gcccaagacc 60
gcctgcacca actgctactg caagaagtgc tgcttccact gccaggtgtg cttcatcacc 120
aagggcctgg gcatcagcta cggccgcaag aagcggccgc agcgccgc 168
```

<210> 90

<211> 102

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: tat cys22
 SF162 protein

<400> 90

Met	Glu	Pro	Val	Asp	Pro	Arg	Leu	Glu	Pro	Trp	Lys	His	Pro	Gly	Ser
1															
															15

Gln	Pro	Lys	Thr	Ala	Gly	Thr	Asn	Cys	Tyr	Cys	Lys	Lys	Cys	Cys	Phe
															30

His	Cys	Gln	Val	Cys	Phe	Ile	Thr	Lys	Gly	Leu	Gly	Ile	Ser	Tyr	Gly
															45

Arg	Lys	Lys	Arg	Arg	Gln	Arg	Arg	Ala	Pro	Pro	Asp	Ser	Glu	Val	
															50

His	Gln	Val	Ser	Leu	Pro	Lys	Gln	Pro	Ala	Ser	Gln	Pro	Gln	Gly	Asp
															65

Pro	Thr	Gly	Pro	Lys	Glu	Ser	Lys	Lys	Val	Glu	Arg	Glu	Thr	Glu	
															85

Thr	Asp	Pro	Val	His	Glx
					100